

# A Posteriori Error Estimation of $hp$ -dG Finite Element Methods for Highly Indefinite Helmholtz Problems (extended version)\*

S. Sauter<sup>†</sup>J. Zech<sup>‡</sup>

## Abstract

In this paper, we will consider an  $hp$ -finite elements discretization of a highly indefinite Helmholtz problem by some dG formulation which is based on the *ultra-weak variational formulation* by Cessenat and Deprés.

We will introduce an a posteriori error estimator and derive reliability and efficiency estimates which are explicit with respect to the wavenumber and the discretization parameters  $h$  and  $p$ . In contrast to the conventional conforming finite element method for indefinite problems, the dG formulation is unconditionally stable and the adaptive discretization process may start from a very coarse initial mesh.

Numerical experiments will illustrate the efficiency and robustness of the method.

*AMS Subject Classifications:* 35J05, 65N12, 65N30

*Key words:* Helmholtz equation at high wavenumber,  $hp$ -finite elements, a posteriori error estimation, discontinuous Galerkin methods, ultra-weak variational formulation

## 1 Introduction

High frequency scattering problems are ubiquitous in many fields of science and engineering and their reliable and efficient numerical simulation pervades numerous engineering applications such as detection (e.g., radar), communication (e.g., wireless), and medicine (e.g., sonic imaging) ([32], [1]). These phenomena are governed by systems of linear partial differential equations (PDEs); the wave equation for elastic waves and the Maxwell equations for electromagnetic scattering. We are here interested in time-harmonic problems where the equation can be reduced to purely spatial problems; for high frequencies these PDEs become highly

---

\*This paper is based on the master's thesis [47], which has been worked out during a visit of the second author at the Institut für Mathematik, Universität Zürich.

<sup>†</sup>Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, e-mail: stas@math.uzh.ch

<sup>‡</sup>Seminar for Applied Mathematics, ETH Zürich, CH-8092 Zürich, Switzerland, e-mail: jakob.zech@sam.math.ethz.ch

indefinite and the development of accurate numerical solution methods is far from being in a mature state.

In this paper we will consider the Helmholtz problem with high wavenumber as our model problem. Although the continuous problem with appropriate boundary conditions has a unique solution, conventional  $hp$ -finite element methods require a minimal resolution condition such that existence and uniqueness is guaranteed on the discrete level (see, e.g., [30], [29], [37], [38], [11]). However, this condition, typically, contains a generic constant  $C$  which is either unknown for specific problems or only very pessimistic estimates are available. This is one of the major motivations for the development of *stabilized* formulations such that the discrete system is always solvable – well-known examples include *least square techniques* [22–24, 40] and discontinuous Galerkin (dG) methods [18–20, 46, 48]. These formulations lead to discrete systems which are unconditionally stable, i.e., no resolution condition is required. Although convergence starts for these methods only after a resolution condition is reached, the stability of the discrete system is considerably improved. The *Ultra Weak Variational Formulation* (UWVF) of Cessenat and Després [9, 10, 13] can be understood as a dG-method that permits the use of non-standard, discontinuous local discretization spaces such as plane waves (see [8, 21, 25, 28]). In this paper we will employ a  $hp$ -dG-finite element method based on the UWVF which was developed in [21] and generalized in [36].

Our focus here is on the development of an a posteriori error estimator for this formulation and its analysis which is explicit with respect to the discretization parameters  $h$ ,  $p$ , and the wavenumber. In contrast to definite elliptic problems, there exist only relatively few publications in the literature on a posteriori estimation for highly indefinite problems (cf. [31], [3], [4], [43], [16]). The papers which are closely related to our work are [26] and [16]: a) In [26], an a posteriori error estimator for the Helmholtz problem has been developed for the interior penalty discontinuous Galerkin (IPDG) method and reliability, efficiency, *and* convergence of the resulting adaptive method is proved. In contrast, we do not prove the convergence of the resulting adaptive method for our dG-formulation. On the other hand, our estimators are properly weighted with the polynomial degree and the estimates are explicit with respect to the wavenumber  $k$ , the mesh width  $h$ , *and* the polynomial degree  $p$ . In addition, the dependence of the constants in the estimates on the wavenumber  $k$  are milder in our approach compared to [26]; b) In [16], a *residual* a posteriori error estimator (cf. [5], [6], [2], [45]) has been developed for the conventional  $hp$ -finite element method. Although efficiency and reliability estimates have been proved, a strict minimal resolution condition is required for the initial finite element space and this is a severe drawback in the context of adaptive discretization.

We will prove in this paper, that our a posteriori error estimator for the  $hp$ -dG-finite element method does not require this strict condition and allows to start the adaptive discretization process from very coarse finite element meshes and no a priori information is required.

The paper is organized as follows. In Section 2, we will introduce the model problem and its dG-discretization by  $hp$ -finite elements. We will recall its unconditional stability and state the quasi-optimal convergence.

Section 3 is devoted to the definition of the residual a posteriori error estimator and we will prove its reliability and efficiency.

In Section 4 we will present an adaptive discretization process and report on numerical experiments which illustrate the behavior of the method for specific model problems such as smooth problems, problems with singularities, problems with constant, varying, and discontinuous wavenumber, and the dependence on the polynomial degree of approximation.

The proof of reliability employs a new  $hp$ - $C^1$  Clément-type interpolation operator which will be defined in Appendix A and  $hp$ -explicit approximation results are proved.

## 2 Discontinuous Galerkin (dG)-Discretization

### 2.1 Helmholtz Equation with Robin Boundary Conditions

Let  $\Omega \subset \mathbb{R}^2$  be a bounded Lipschitz domain with boundary  $\partial\Omega$ . The scalar product in  $L^2(\Omega)$  is denoted by  $(u, v) := \int_{\Omega} u \bar{v}$  and the norm by  $\|\cdot\|$ .

For  $s > 0$ , the space  $H^s(\Omega)$  is the usual Sobolev space with norm  $\|\cdot\|_{H^s(\Omega)}$ . The dual space is denoted by  $(H^s(\Omega))'$  and the trace spaces by  $H^\sigma(\partial\Omega)$  with norm  $\|\cdot\|_{H^\sigma(\partial\Omega)}$ . For  $\sigma = 0$ , we write  $\|\cdot\|_{\partial\Omega}$  short for  $\|\cdot\|_{L^2(\partial\Omega)}$ . The seminorms containing only the highest derivatives are denoted by  $|\cdot|_{H^s(\Omega)}$  and  $|\cdot|_{H^\sigma(\partial\Omega)}$ .

For given  $f \in L^2(\Omega)$ ,  $g \in L^2(\partial\Omega)$  we consider the Helmholtz equation with Robin boundary condition

$$\begin{aligned} -\Delta u - k^2 u &= f \quad \text{in } \Omega, \\ \partial_{\mathbf{n}} u + iku &= g \quad \text{on } \partial\Omega, \end{aligned}$$

where  $\partial_{\mathbf{n}} u$  denotes the outer normal derivative of  $u$  on the boundary. In most parts of this paper we assume that  $k$  is a positive constant. This is a simplification compared to the following more general case: There exist positive constants  $\kappa$  and  $k_{\max}$  such that

$$\begin{aligned} k &\in L^\infty(\Omega, \mathbb{R}), \quad 1 < \kappa \leq k(x) \leq k_{\max} < \infty, \\ k &= \kappa \quad \text{in a neighborhood of } \partial\Omega. \end{aligned} \tag{2.1}$$

We define the method for, possibly, variable wavenumbers  $k$  which satisfy (2.1) while the error analysis is restricted to the constant case. In the section on numerical experiments, we will again consider variable wavenumbers  $k$ .

The weak formulation reads: Find  $u \in H^1(\Omega)$  such that

$$a(u, v) = F(v) \quad \forall v \in H^1(\Omega) \tag{2.2a}$$

with the sesquilinear form  $a : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{C}$  and linear form  $F : H^1(\Omega) \rightarrow \mathbb{C}$  defined by

$$a(u, v) := \int_{\Omega} (\langle \nabla u, \overline{\nabla v} \rangle - k^2 u \bar{v}) + i \int_{\partial\Omega} k u \bar{v} \quad \text{and} \quad F(v) := \int_{\Omega} f \bar{v} + \int_{\partial\Omega} g \bar{v}. \tag{2.2b}$$

The assumptions on the data can be weakened to  $f \in (H^1(\Omega))'$  and  $g \in H^{-1/2}(\partial\Omega)$ . In this case the integrals in (2.2b) are understood as dual pairings.

It is well-known that this problem has a unique solution which depends continuously on the data.

**Definition 2.1.** Let  $k$  satisfy (2.1). On  $H^1(\Omega)$ , we introduce the norm

$$\|u\|_{\mathcal{H}} := \|\nabla u\| + \|ku\|.$$

**Theorem 2.2.** Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded Lipschitz domain and let  $k = \kappa > 1$  be constant.

- a. There exists a constant  $C(\Omega, \kappa) > 0$  such that for every  $f \in (H^1(\Omega))'$  and  $g \in H^{-1/2}(\partial\Omega)$ , there exists a unique solution  $u \in H^1(\Omega)$  of problem (2.2) which satisfies

$$\|u\|_{\mathcal{H}} \leq C(\Omega, \kappa) \left( \|F\|_{(H^1(\Omega))'} + \|g\|_{H^{-1/2}(\partial\Omega)} \right).$$

- b. Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded star-shaped domain with smooth boundary or a bounded convex domain. There exists a constant  $C(\Omega) > 0$  (depending only on  $\Omega$ ) such that for any  $f \in L^2(\Omega)$ ,  $g \in H^{1/2}(\partial\Omega)$ , the solution of (2.2) satisfies

$$\begin{aligned} \|u\|_{\mathcal{H}} &\leq C(\Omega) (\|f\| + \|g\|_{\partial\Omega}), \\ \|u\|_{H^2(\Omega)} &\leq C(\Omega) (1 + \kappa) \left( \|f\| + \|g\|_{\partial\Omega} + \|g\|_{H^{1/2}(\partial\Omega)} \right). \end{aligned}$$

For a proof we refer to [34, Prop. 8.1.3 and .4].

**Remark 2.3.** Let  $\Omega \subseteq \mathbb{R}^2$  be a polygonal Lipschitz domain and let  $k = \kappa > 1$  be a constant. For  $f \in L^2(\Omega)$  and  $g \in H_{\text{pw}}^{1/2}(\partial\Omega) := \{g \in L^2(\partial\Omega) : g \text{ is edgewise in } H^{1/2}\}$ , the classical elliptic regularity theory shows that the unique solution  $u$  of (2.2) is in  $H^{3/2+\varepsilon}(\Omega)$  for some  $\varepsilon > 0$  depending on  $\Omega$  and we briefly sketch the argument: We write (2.2) in the following strong form

$$\begin{aligned} -\Delta u &= \tilde{f} := f + k^2 u \quad \text{in } \Omega, \\ \partial_{\mathbf{n}} u &= \tilde{g} := g - i k u \quad \text{on } \partial\Omega. \end{aligned}$$

Since the solution  $u$  of (2.2) is in  $H^1(\Omega)$ , we have  $\tilde{f} \in L^2(\Omega)$  and  $\tilde{g} \in H_{\text{pw}}^{1/2}(\partial\Omega)$ . From [38, Lemma A1], we conclude that there exists a lifting operator  $\mathcal{L} : H_{\text{pw}}^{1/2}(\partial\Omega) \rightarrow H^2(\Omega)$  such that  $G := \mathcal{L}(g)$  satisfies  $\partial_{\mathbf{n}} G = g$  and  $\|G\|_{H^2(\Omega)} \leq C \|g\|_{H_{\text{pw}}^{1/2}(\partial\Omega)}$ . Thus, the ansatz  $u = u_0 + \tilde{G}$  with  $\tilde{G} := \mathcal{L}(\tilde{g})$  leads to

$$\begin{aligned} -\Delta u_0 &= \check{f} := \tilde{f} + \Delta \tilde{G} \quad \text{in } \Omega, \\ \partial_{\mathbf{n}} u_0 &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

with  $\check{f} \in L^2(\Omega)$ . From [33, (7.22)] we obtain that the solution  $u_0$ , and thus also  $u$ , then is in  $H^{3/2+\varepsilon}(\Omega)$  for some  $\varepsilon > 0$ .

## 2.2 hp-Finite Elements

Let  $\Omega \subset \mathbb{R}^2$  be a polygonal domain and let  $\mathcal{T} := \{K_i : 1 \leq i \leq N\}$  denote a simplicial finite element mesh which is *conforming* in the sense that there are no *hanging* nodes. With each element  $K \in \mathcal{T}$  we associate a polynomial degree  $p_K \in \mathbb{N}_{\geq 1}$ .

The diameter of an element  $K \in \mathcal{T}$  is denoted by  $h_K := \text{diam } K$  and the maximal mesh width is  $h_{\mathcal{T}} := \max \{h_K : K \in \mathcal{T}\}$ . The *minimal* polynomial degree is

$$p_{\mathcal{T}} := \min \{p_K : K \in \mathcal{T}\}.$$

The *shape regularity* of  $\mathcal{T}$  is described by the constant

$$\rho_{\mathcal{T}} := \max \left\{ \frac{h_K}{\text{diam } \mathcal{B}_K} : K \in \mathcal{T} \right\}, \quad (2.3)$$

where  $\mathcal{B}_K$  is the maximal inscribed ball in  $K$ . Since  $\mathcal{T}$  contains finitely many simplices, the constant  $\rho_{\mathcal{T}}$  is always bounded but becomes large if the simplices are degenerate, e.g., are flat or needle-shaped. The constants in the following estimates depend on the mesh via the constant  $\rho_{\mathcal{T}}$ ; they are bounded for any fixed  $\rho_{\mathcal{T}}$  but, possibly, become large for large  $\rho_{\mathcal{T}}$ .

Concerning the polynomial degree distribution we assume throughout the paper that the polynomial degrees of neighboring elements are comparable<sup>1</sup>:

$$\rho_{\mathcal{T}}^{-1} (p_K + 1) \leq p_{K'} + 1 \leq \rho_{\mathcal{T}} (p_K + 1) \quad \forall K, K' \in \mathcal{T} \text{ with } K \cap K' \neq \emptyset. \quad (2.4)$$

By convention the triangles  $K \in \mathcal{T}$  are closed sets. The boundary of a triangle  $K \in \mathcal{T}$  consists of three one-dimensional (relatively closed) edges which are collected in the set  $\mathcal{E}(K)$ . The subset  $\mathcal{E}^I(K) \subseteq \mathcal{E}(K)$  of *inner* edges consists of all edges  $e \in \mathcal{E}(K)$  whose relative interior lie in (the open set)  $\Omega$  while  $\mathcal{E}^B(K) := \mathcal{E}(K) \setminus \mathcal{E}^I(K)$  is the set of *boundary* edges. Further we set

$$\partial^B K := \partial K \cap \partial \Omega \quad \text{and} \quad \partial^I K := \partial K \setminus \partial^B K.$$

The conformity of the mesh implies that any  $e \in \mathcal{E}^I(K)$  is shared by two and only two triangles in  $\mathcal{T}$ . The sets of inner/boundary/all edges  $\mathcal{E}^I$ ,  $\mathcal{E}^B$ ,  $\mathcal{E}$ , are defined by

$$\mathcal{E}^I := \{e \in \mathcal{E}^I(K) : K \in \mathcal{T}\}, \quad \mathcal{E}^B := \{e \in \mathcal{E}^B(K) : K \in \mathcal{T}\}, \quad \mathcal{E} := \mathcal{E}^I \cup \mathcal{E}^B.$$

The interior skeleton  $\mathfrak{S}^I$  is given by

$$\mathfrak{S}^I := \bigcup_{K \in \mathcal{T}} \partial^I K.$$

Next we introduce patches associated with an edge  $e$  or an element  $K$  of the triangulation

$$\omega_e := \bigcup_{\{K' \in \mathcal{T} : e \cap K' \neq \emptyset\}} K' \quad \text{and} \quad \omega_K := \bigcup_{\{K' \in \mathcal{T} : K \cap K' \neq \emptyset\}} K'.$$

Furthermore, we employ the notation

$$p_e := \min_{\substack{K \in \mathcal{T} \\ e \subset \partial K}} p_K \quad \text{and} \quad h_e := |e| \quad \text{with the length } |e| \text{ of } e. \quad (2.5)$$

---

<sup>1</sup>We use here the same constant  $\rho_{\mathcal{T}}$  as for the shape regularity to simplify the notation.

We define the *mesh functions*  $\mathfrak{h}_{\mathcal{T}}, \mathfrak{p}_{\mathcal{T}} \in L^\infty(\Omega)$  and  $\mathfrak{h}_{\mathcal{E}}, \mathfrak{p}_{\mathcal{E}} \in L^\infty(\mathfrak{S})$  by

$$\forall K \in \mathcal{T} : \quad (\mathfrak{h}_{\mathcal{T}})|_K := h_K, \quad \mathfrak{p}_{\mathcal{T}}|_K := p_K \quad \text{and} \quad \forall e \in \mathcal{E} : \quad (\mathfrak{h}_{\mathcal{E}})|_e := h_e, \quad \mathfrak{p}_{\mathcal{E}}|_e := p_e.$$

We skip the indices  $\mathcal{T}$  and  $\mathcal{E}$  and write short  $\mathfrak{h}, \mathfrak{p}$  if no confusion is possible. In the error estimates, the quantity  $k\mathfrak{h}/\mathfrak{p}$  will play an important role since it is a measure how well the  $hp$ -finite element space *resolves* the oscillations in the solution. Therefore we define

$$M_{\frac{k\mathfrak{h}}{\mathfrak{p}}} := \max \left\{ \left\| \frac{\mathfrak{h}}{\mathfrak{p}} k \right\|_{L^\infty(\mathfrak{S})}, \left\| \frac{\mathfrak{h}}{\mathfrak{p}} k \right\|_{L^\infty(\Omega)} \right\}. \quad (2.6)$$

The non-conforming  $hp$ -finite element space for the mesh  $\mathcal{T}$  with local polynomials of degree  $p_K$  is given by

$$S_{\mathcal{T}}^{\mathfrak{p}} := \{u \in L^2(\Omega) : u|_K \in \mathbb{P}_{p_K} \quad \forall K \in \mathcal{T}\}. \quad (2.7)$$

Here  $\mathbb{P}_p$  denotes the space of bivariate polynomials of maximal total degree  $p$ . For a subset  $\omega \subset \Omega$ , we write  $\mathbb{P}_p(\omega)$  to indicate explicitly that we consider  $u \in \mathbb{P}_p(\omega)$  as a polynomial on  $\omega$ .

Finally, throughout this paper  $C > 0$  stands for a generic constant that does not depend on the parameters  $k, h_K$ , and  $p_K$  and may change its value in each occurrence.

## 2.3 dG Formulation

For the discretization of the Helmholtz problem we employ a dG formulation which has been derived from the ultra-weak variational formulation (cf. [9,10,13]) in [21], [25], and generalized in [36]. It involves jumps and mean values across edges which we will introduce next. For an inner edge  $e \in \mathcal{E}^I$  with two adjacent triangles  $K, K' \in \mathcal{T}$  we set for simplexwise sufficiently smooth functions  $v$  and vector valued functions  $\mathbf{w}$

$$\begin{aligned} \llbracket v \rrbracket|_e &:= (v|_K)|_e - (v|_{K'})|_e, & \{v\}|_e &:= \frac{1}{2}((v|_K)|_e + (v|_{K'})|_e), \\ \llbracket v \rrbracket_N|_e &:= (v|_K)|_e \mathbf{n}_K + (v|_{K'})|_e \mathbf{n}_{K'}, & \llbracket \mathbf{w} \rrbracket_N|_e &:= (\mathbf{w}|_K)|_e \cdot \mathbf{n}_K + (\mathbf{w}|_{K'})|_e \cdot \mathbf{n}_{K'}, \end{aligned}$$

where  $\mathbf{n}_K, \mathbf{n}_{K'}$  are the respective outer normal vectors on the boundary of  $K$  and  $K'$  and “ $\cdot$ ” denotes the Euclidean scalar product. The sign in  $\llbracket v \rrbracket|_e$  is arbitrary.

The dG-discretization of (2.2) reads: Find  $u_{\mathcal{T}} \in S_{\mathcal{T}}^{\mathfrak{p}}$  such that

$$a_{\mathcal{T}}(u_{\mathcal{T}}, v) = F_{\mathcal{T}}(v) \quad \forall v \in S_{\mathcal{T}}^{\mathfrak{p}} \quad (2.8a)$$

with the sesquilinear form

$$\begin{aligned} a_{\mathcal{T}}(u, v) &:= (\nabla_{\mathcal{T}} u, \nabla_{\mathcal{T}} v) - k^2(u, v) - (\llbracket u \rrbracket_N, \{\nabla_{\mathcal{T}} v\})_{\mathfrak{S}^I} - (\{\nabla_{\mathcal{T}} u\}, \llbracket v \rrbracket_N)_{\mathfrak{S}^I} \\ &\quad - \left( \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} u, \nabla_{\mathcal{T}} v \cdot \mathbf{n} \right)_{\partial\Omega} - \left( \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} \nabla_{\mathcal{T}} u \cdot \mathbf{n}, v \right)_{\partial\Omega} \\ &\quad - \frac{1}{i} \left( \mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}} \llbracket \nabla_{\mathcal{T}} u \rrbracket_N, \llbracket \nabla_{\mathcal{T}} v \rrbracket_N \right)_{\mathfrak{S}^I} - \frac{1}{i} \left( \mathfrak{d} \frac{\mathfrak{h}}{\mathfrak{p}} \nabla_{\mathcal{T}} u \cdot \mathbf{n}, \nabla_{\mathcal{T}} v \cdot \mathbf{n} \right)_{\partial\Omega} \\ &\quad + i \left( \mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \llbracket u \rrbracket_N, \llbracket v \rrbracket_N \right)_{\mathfrak{S}^I} + i \left( k \left( 1 - \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} \right) u, v \right)_{\partial\Omega}, \end{aligned} \quad (2.8b)$$

where  $\nabla_{\mathcal{T}}$  denotes the simplexwise gradient,  $\Delta_{\mathcal{T}}$  the simplexwise Laplacean, and  $(\cdot, \cdot)_{\mathfrak{S}^I}$ ,  $(\cdot, \cdot)_{\partial\Omega}$  are the  $L^2(\mathfrak{S}^I)$  and  $L^2(\partial\Omega)$  scalar products. Moreover, the fixed constants

$$\mathfrak{a} > 0, \quad \mathfrak{b} > 0, \quad \mathfrak{d} > 0$$

are at our disposal and will be adjusted later. The functional  $F_{\mathcal{T}}$  is defined by

$$F_{\mathcal{T}}(v) := (f, v) - \left( \frac{\mathfrak{d}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} g, \nabla_{\mathcal{T}} v \cdot \mathbf{n} \right)_{\partial\Omega} + \left( \left( 1 - \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} \right) g, v \right)_{\partial\Omega}. \quad (2.8c)$$

**Remark 2.4.** In [36, Section 3, Remark 3.2] it is proved that the condition:

$$\left\| \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} \right\|_{L^\infty(\partial\Omega)} < 1/2 \quad (2.9)$$

implies the unique solvability of the discrete system (2.8). As a consequence, the discrete system is always solvable for sufficiently small  $\mathfrak{d} > 0$ . In addition, for any fixed  $\mathfrak{d} > 0$ , condition (2.9) can be regarded as an explicit condition on  $\mathfrak{h}$  and  $\mathfrak{p}$ . This is a significant improvement compared to the condition

$$\left\| \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} \right\|_{L^\infty(\partial\Omega)} < C \quad \text{for “sufficiently” small } C > 0$$

which is typically imposed for the solvability of the standard finite element discretization of the Helmholtz problem (cf. [29, Sec. 4.1.3] and [37, 38]).

**Remark 2.5.** For  $s > 0$ , let the broken Sobolev space  $H_{\mathcal{T}}^s(\Omega)$  be defined by

$$H_{\mathcal{T}}^s(\Omega) := \{u \in L^2(\Omega) \mid \forall K \in \mathcal{T} : u|_K \in H^s(K)\}.$$

Then,  $a_{\mathcal{T}}(\cdot, \cdot)$  can be extended to a sesquilinear form on  $H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega) \times H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega)$  and  $F_{\mathcal{T}}(\cdot)$  to a linear functional  $F_{\mathcal{T}} : H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega) \rightarrow \mathbb{C}$  for any  $\varepsilon > 0$ .

## 2.4 Discrete Stability and Convergence

The following mesh-dependent norms on  $H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega)$  for  $\varepsilon > 0$  have been introduced in [21]:

$$\|v\|_{\text{dG}} := \left( \|\nabla_{\mathcal{T}} v\|^2 + \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} [\nabla_{\mathcal{T}} v]_N \right\|_{\mathfrak{S}^I}^2 + \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} [v]_N \right\|_{\mathfrak{S}^I}^2 \right. \quad (2.10a_1)$$

$$\left. + \left\| \sqrt{\mathfrak{d} \frac{\mathfrak{h}}{\mathfrak{p}}} \nabla_{\mathcal{T}} v \cdot \mathbf{n} \right\|_{\partial\Omega}^2 + \left\| \sqrt{k \left( 1 - \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} \right)} v \right\|_{\partial\Omega}^2 + \|kv\|^2 \right)^{1/2}, \quad (2.10a_2)$$

$$\|v\|_{\text{dG}^+} := \left( \|v\|_{\text{dG}}^2 + \left\| \left( \mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \right)^{-1/2} \{\nabla_{\mathcal{T}} v\} \right\|_{\mathfrak{S}^I}^2 \right)^{1/2}. \quad (2.10b)$$

Before formulating the stability and convergence theorem, we have to introduce some notation.

The *adjoint Helmholtz problem* reads: For given  $w \in L^2(\Omega)$ , find  $z \in H^1(\Omega)$  such that

$$a(v, z) = (v, w) \quad \forall v \in H^1(\Omega). \quad (2.11)$$

The assumptions of Theorem 2.2 ensure well-posedness of the adjoint problem (cf. [34, Prop. 8.1.4], [12], [17, Thm. 2.4], [36]) and defines a bounded solution operator  $Q_k^* : L^2(\Omega) \rightarrow H^1(\Omega)$ ,  $w \mapsto z$ .

**Lemma 2.6.** *Let  $\Omega \subset \mathbb{R}^2$  be a polygonal Lipschitz domain and let  $w \in L^2(\Omega)$ . Then, (2.11) is a well-posed problem. Denote its solution by  $z$ . Then  $z$  satisfies  $z \in H^{3/2+\varepsilon}$  for some  $\varepsilon > 0$  depending on  $\Omega$  and moreover*

$$a_{\mathcal{T}}(v, z) = (v, w) \quad \forall v \in H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega).$$

This follows from [36, Rem. 2.6, Lem. 2.7.].

The key role for the convergence estimates for Helmholtz-type problems is played by the *adjoint approximation property* which will be defined next.

**Definition 2.7.** *Let  $S \subset H^1(\Omega)$  be a subspace of  $H^1(\Omega)$ . Then the adjoint approximation property is given by*

$$\sigma_k^*(S) := \sup_{g \in L^2(\Omega) \setminus \{0\}} \inf_{v \in S} \frac{\|Q_k^*(k^2 g) - v\|_{\text{dG}^+}}{\|kg\|}. \quad (2.12)$$

There holds the following result on uniqueness and quasi-optimality of the dG-finite element solution (see [36, Sec. 3], [47, Rem. 2.3.1, .2 and Thm. 2.3.5], and Remark 2.3).

**Theorem 2.8.** *Let  $k = \kappa$  be constant satisfying (2.1). Let  $\Omega \subset \mathbb{R}^2$  be a polygonal Lipschitz domain. Furthermore assume that the constant  $\mathfrak{a}$  in (2.8b) is chosen sufficiently large and condition (2.9) is fulfilled. Then, the dG-problem (2.8) has a unique solution  $u_{\mathcal{T}} \in S_{\mathcal{T}}^{\mathfrak{p}}$ . If, in addition, the adjoint approximation condition*

$$\sigma_k^*(S_{\mathcal{T}}^{\mathfrak{p}}) \leq C_* \quad (2.13)$$

*holds for some  $C_* > 0$ , then, the quasi-optimal error estimate*

$$\|u - u_{\mathcal{T}}\|_{\text{dG}} \leq C \inf_{v \in S_{\mathcal{T}}^{\mathfrak{p}}} \|u - v\|_{\text{dG}^+}$$

*holds, where  $C$  is independent of  $k$ ,  $\mathfrak{h}$ , and  $\mathfrak{p}$ .*

### 3 A Posteriori Error Estimation

In this section we will derive and analyze a residual type a posteriori estimator for the dG-formulation (2.8) of the Helmholtz problem (2.2). General techniques of a posteriori error estimation for elliptic problems are described in [2], [39], [45] while the focus in [15] is on



dG-methods. A posteriori error estimation for the conventional conforming discretization of the Helmholtz problem are described in [16] and for an IPDG method in [26].

For the derivation of an a posteriori error estimator for the dG-formulation of the Helmholtz problem the main challenges are a) the lower order term  $-k^2(\cdot, \cdot)$  in the sesquilinear forms  $a(\cdot, \cdot)$  and  $a_{\mathcal{T}}(\cdot, \cdot)$ , which causes the problem to be highly indefinite and b) the integrals in (2.8b) containing the mean of the gradient on interior edges, which have the effect that  $a_{\mathcal{T}}(\cdot, \cdot) + 2k^2(\cdot, \cdot)_{L^2}$  is not coercive on  $H_{\mathcal{T}}^{3/2+\varepsilon} \cap H^1(\Omega)$ ,  $\varepsilon > 0$ , with respect to the norm  $\|\cdot\|_{\mathcal{H}}$ .

### 3.1 The Residual Error Estimator

**Definition 3.1.** For  $v \in S_{\mathcal{T}}^{\mathfrak{p}}$  and  $K \in \mathcal{T}$ , the local error estimator is

$$\eta_K(v) := (\eta_{R_K}^2(v) + \eta_{E_K}^2(v) + \eta_{J_K}^2(v))^{1/2} \quad (3.1a)$$

with the internal residual  $\eta_{R_K}$ , the edge residual  $\eta_{E_K}$ , and the trace residual  $\eta_{J_K}$  given by

$$\eta_{R_K}(v) := \left( \frac{h_K}{p_K} \right) \|\Delta_{\mathcal{T}} v + k^2 v + f\|_{L^2(K)} \quad (3.1b)$$

$$\eta_{E_K}(v) := \left\{ \frac{1}{2} \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} v \rrbracket_N \right\|_{\partial^I K}^2 + \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} v - \mathfrak{i} k v) \right\|_{\partial^B K}^2 \right\}^{1/2}, \quad (3.1c)$$

$$\eta_{J_K}(v) := \frac{1}{\sqrt{2}} \left\| \sqrt{\frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket v \rrbracket \right\|_{\partial^I K}. \quad (3.1d)$$

The global error estimator is

$$\eta(v) := (\eta_R^2(v) + \eta_E^2(v) + \eta_J^2(v))^{1/2} \quad (3.2a)$$

with

$$\eta_R(v) := \left( \sum_{K \in \mathcal{T}} \eta_{R_K}^2(v) \right)^{1/2}, \quad \eta_E(v) := \left( \sum_{K \in \mathcal{T}} \eta_{E_K}^2(v) \right)^{1/2}, \quad \eta_J(v) := \left( \sum_{K \in \mathcal{T}} \eta_{J_K}^2(v) \right)^{1/2}. \quad (3.2b)$$

For the solution  $u_{\mathcal{T}}$  of (2.8), we write  $\eta$  short for  $\eta(u_{\mathcal{T}})$  and similarly for  $\eta_{R_K}$ ,  $\eta_{E_K}$ , etc.

### 3.2 Reliability

We start the derivation of the reliability estimate by bounding the dG-norm of the error by parts of the estimator plus the  $k$ -weighted  $L^2$ -norm of the error.

**Lemma 3.2.** Let  $\Omega \subseteq \mathbb{R}^2$  be a polygonal Lipschitz domain. Let  $k = \kappa$  be constant satisfying (2.1) and let  $p_{\mathcal{T}} \geq 1$ . Let  $u \in H^{3/2+\varepsilon}(\Omega)$  be the solution of (2.2) for some  $\varepsilon > 0$  and assume that  $u_{\mathcal{T}} \in S_{\mathcal{T}}^{\mathfrak{p}}$  solves (2.8). Furthermore assume that the constant  $\mathfrak{a}$  in (2.8b) is chosen

sufficiently large. Then, there exists a constant  $C > 0$  which only depends on  $\mathfrak{b}$ ,  $\mathfrak{d}$ ,  $\rho_{\mathcal{T}}$ , and  $\Omega$  such that

$$\|u - u_{\mathcal{T}}\|_{\text{dG}} \leq C \left( C_{\text{conf}}^{3/2} \eta(u_{\mathcal{T}}) + C_{\text{conf}}^{1/2} \|k(u - u_{\mathcal{T}})\| \right).$$

where

$$C_{\text{conf}} := 1 + M_{\frac{\mathfrak{kh}}{\mathfrak{p}}}.$$

Before we prove this lemma we compute an alternative representation of the term  $a_{\mathcal{T}}(u - u_{\mathcal{T}}, v)$  which will be used frequently in the following.

**Lemma 3.3.** *Let  $u \in H^{3/2+\varepsilon}(\Omega)$  be the solution of (2.2) for some  $\varepsilon > 0$  and assume that  $u_{\mathcal{T}} \in S_{\mathcal{T}}^{\mathfrak{p}}$  solves (2.8). Then, we have for  $v \in H_{\mathcal{T}}^{3/2+\tilde{\varepsilon}}(\Omega)$ ,  $\tilde{\varepsilon} > 0$ ,*

$$\begin{aligned} a_{\mathcal{T}}(u - u_{\mathcal{T}}, v) &= (f + \Delta_{\mathcal{T}}u_{\mathcal{T}} + k^2u_{\mathcal{T}}, v) - (\llbracket \nabla_{\mathcal{T}}u_{\mathcal{T}} \rrbracket_N, \{v\})_{\mathfrak{S}^I} + (\llbracket u_{\mathcal{T}} \rrbracket_N, \{\nabla_{\mathcal{T}}v\})_{\mathfrak{S}^I} \\ &\quad + \left( \left(1 - \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}}\right) (g - \partial_{\mathbf{n}}u_{\mathcal{T}} - \mathfrak{i}ku_{\mathcal{T}}), v \right)_{\partial\Omega} - \left( \frac{\mathfrak{d}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} (g - \partial_{\mathbf{n}}u_{\mathcal{T}} - \mathfrak{i}ku_{\mathcal{T}}), \partial_{\mathbf{n}}v \right)_{\partial\Omega} \\ &\quad - \left( \mathfrak{i}\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \llbracket u_{\mathcal{T}} \rrbracket_N, \llbracket v \rrbracket_N \right)_{\mathfrak{S}^I} + \left( \frac{\mathfrak{b}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} \llbracket \nabla_{\mathcal{T}}u_{\mathcal{T}} \rrbracket_N, \llbracket \nabla_{\mathcal{T}}v \rrbracket_N \right)_{\mathfrak{S}^I}. \end{aligned} \quad (3.3)$$

*Proof.* Note that  $-\Delta u - k^2u = f$  in  $\Omega$ . Integrating by parts we obtain with the “dG-magic formula”

$$\begin{aligned} (\nabla_{\mathcal{T}}(u - u_{\mathcal{T}}), \nabla_{\mathcal{T}}v) - (k^2(u - u_{\mathcal{T}}), v) &= (f + \Delta_{\mathcal{T}}u_{\mathcal{T}} + k^2u_{\mathcal{T}}, v) \\ &\quad + (\nabla_{\mathcal{T}}(u - u_{\mathcal{T}}) \cdot \mathbf{n}, v)_{\partial\Omega} + (\llbracket \nabla_{\mathcal{T}}(u - u_{\mathcal{T}}) \rrbracket_N, \{v\})_{\mathfrak{S}^I} + (\{\nabla_{\mathcal{T}}(u - u_{\mathcal{T}})\}, \llbracket v \rrbracket_N)_{\mathfrak{S}^I}. \end{aligned}$$

By inserting this into (2.8b) and using  $\partial_{\mathbf{n}}(u - u_{\mathcal{T}}) + \mathfrak{i}k(u - u_{\mathcal{T}}) = g - \partial_{\mathbf{n}}u_{\mathcal{T}} - \mathfrak{i}ku_{\mathcal{T}}$  on  $\partial\Omega$  we get

$$\begin{aligned} a_{\mathcal{T}}(u - u_{\mathcal{T}}, v) &= (f + \Delta_{\mathcal{T}}u_{\mathcal{T}} + k^2u_{\mathcal{T}}, v) \\ &\quad + (\llbracket \nabla_{\mathcal{T}}(u - u_{\mathcal{T}}) \rrbracket_N, \{v\})_{\mathfrak{S}^I} - (\llbracket u - u_{\mathcal{T}} \rrbracket_N, \{\nabla_{\mathcal{T}}v\})_{\mathfrak{S}^I} \\ &\quad - \left( \frac{\mathfrak{b}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} \llbracket \nabla_{\mathcal{T}}(u - u_{\mathcal{T}}) \rrbracket_N, \llbracket \nabla_{\mathcal{T}}v \rrbracket_N \right)_{\mathfrak{S}^I} + \left( \mathfrak{i}\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \llbracket u - u_{\mathcal{T}} \rrbracket_N, \llbracket v \rrbracket_N \right)_{\mathfrak{S}^I} \\ &\quad + \left( \left(1 - \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}}\right) (g - \partial_{\mathbf{n}}u_{\mathcal{T}} - \mathfrak{i}ku_{\mathcal{T}}), v \right)_{\partial\Omega} - \left( \frac{\mathfrak{d}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} (g - \partial_{\mathbf{n}}u_{\mathcal{T}} - \mathfrak{i}ku_{\mathcal{T}}), \nabla_{\mathcal{T}}v \cdot \mathbf{n} \right)_{\partial\Omega}. \end{aligned}$$

The regularity of the solution  $u \in H^{3/2+\varepsilon}(\Omega)$  for some  $\varepsilon > 0$  implies that all internal jumps of  $u$  vanish and (3.3) follows.  $\square$

*Proof. (Lemma 3.2).* We first assume  $p_{\mathcal{T}} \geq 5$ .

**Part 1.** We introduce the sesquilinear form  $\tilde{a}_{\mathcal{T}} : H_{\mathcal{T}}^1(\Omega) \times H_{\mathcal{T}}^1(\Omega) \rightarrow \mathbb{C}$  by

$$\tilde{a}_{\mathcal{T}}(v, w) := (\nabla_{\mathcal{T}}v, \nabla_{\mathcal{T}}w) + (k^2v, w) + \mathfrak{i}(kv, w)_{\partial\Omega}$$

and the associated norm

$$\|v\|_{\tilde{a}} := \sqrt{|\tilde{a}_{\mathcal{T}}(v, v)|}.$$

In Part 2, we will prove

$$\|u - u_{\mathcal{T}}\|_{\tilde{a}} \leq CC_{\text{conf}} (\eta_R^2 + \eta_J^2 + \eta_E^2)^{1/2} + 2 \|k(u - u_{\mathcal{T}})\|. \quad (3.4)$$

The combination of

$$\frac{1}{2} \left( \|\nabla_{\mathcal{T}} v\|^2 + \|kv\|^2 + \|k^{1/2}v\|_{\partial\Omega}^2 \right) \leq \|v\|_{\tilde{a}}^2 \leq \|\nabla_{\mathcal{T}} v\|^2 + \|kv\|^2 + \|k^{1/2}v\|_{\partial\Omega}^2$$

with the definition of the dG-norm leads to

$$\begin{aligned} \|u - u_{\mathcal{T}}\|_{\text{dG}}^2 \leq & 2 \|u - u_{\mathcal{T}}\|_{\tilde{a}}^2 + \left\| \sqrt{\mathfrak{d} \frac{\mathfrak{h}}{\mathfrak{p}}} \nabla_{\mathcal{T}} (u - u_{\mathcal{T}}) \cdot \mathbf{n} \right\|_{\partial\Omega}^2 \\ & + \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} (u - u_{\mathcal{T}}) \rrbracket_N \right\|_{\mathfrak{S}^I}^2 + \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket (u - u_{\mathcal{T}}) \rrbracket_N \right\|_{\mathfrak{S}^I}^2. \end{aligned} \quad (3.5)$$

To estimate the boundary term in (3.5), we employ  $\partial_{\mathbf{n}} u = g - iku$  so that for  $e \in \mathcal{E}^B$  it holds

$$\begin{aligned} \left\| \sqrt{\mathfrak{d} \frac{\mathfrak{h}}{\mathfrak{p}}} \nabla_{\mathcal{T}} (u - u_{\mathcal{T}}) \cdot \mathbf{n} \right\|_e &= \sqrt{\mathfrak{d} \frac{h_e}{p_e}} \|g - \partial_{\mathbf{n}} u_{\mathcal{T}} - iku\|_e \\ &\leq \sqrt{\mathfrak{d} \frac{h_e}{p_e}} \|g - \partial_{\mathbf{n}} u_{\mathcal{T}} - iku_{\mathcal{T}}\|_e + \sqrt{\mathfrak{d} M_{\frac{\mathfrak{kh}}{\mathfrak{p}}}} \|k^{1/2}(u - u_{\mathcal{T}})\|_e. \end{aligned}$$

A summation over all  $e \in \mathcal{E}^B$  leads to

$$\begin{aligned} \left\| \sqrt{\mathfrak{d} \frac{\mathfrak{h}}{\mathfrak{p}}} \nabla_{\mathcal{T}} (u - u_{\mathcal{T}}) \cdot \mathbf{n} \right\|_{\partial\Omega}^2 &\leq 2\mathfrak{d} \left( \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - iku_{\mathcal{T}}) \right\|_{\partial\Omega}^2 \right. \\ &\quad \left. + M_{\frac{\mathfrak{kh}}{\mathfrak{p}}} \|k^{1/2}(u - u_{\mathcal{T}})\|_{\partial\Omega}^2 \right) \\ &\leq \frac{2\mathfrak{d}}{p_{\mathcal{T}}} \eta_E^2 + 2\mathfrak{d} M_{\frac{\mathfrak{kh}}{\mathfrak{p}}} \|u - u_{\mathcal{T}}\|_{\tilde{a}}^2. \end{aligned} \quad (3.6)$$

For the inner jump terms in (3.5) we obtain

$$\left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} (u - u_{\mathcal{T}}) \rrbracket_N \right\|_{\mathfrak{S}^I} = \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \leq \sqrt{2} \eta_E \quad (3.7a)$$

$$\left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u - u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} = \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \leq \sqrt{2} \eta_J, \quad (3.7b)$$

since the regularity assumptions on  $u$  imply that the corresponding jump terms vanish.

The combination of (3.5), (3.4), (3.6), (3.7) yields

$$\begin{aligned}
\|u - u_{\mathcal{T}}\|_{\text{dG}}^2 &\leq \left(2 + 2\mathfrak{d}M_{\frac{\text{kh}}{\mathfrak{p}}}\right) \|u - u_{\mathcal{T}}\|_{\tilde{a}}^2 + \left(2 + \frac{2\mathfrak{d}}{p_{\mathcal{T}}}\right) \eta_E^2 + 2\eta_J^2 \\
&\leq \left(2 + 2\mathfrak{d}M_{\frac{\text{kh}}{\mathfrak{p}}}\right) (2C^2C_{\text{conf}}^2 (\eta_R^2 + \eta_J^2 + \eta_E^2) + 4\|k(u - u_{\mathcal{T}})\|^2) \\
&\quad + \left(2 + \frac{2\mathfrak{d}}{p_{\mathcal{T}}}\right) \eta_E^2 + 2\eta_J^2 \\
&\leq C(C_{\text{conf}}^3 (\eta_R^2 + \eta_J^2 + \eta_E^2) + C_{\text{conf}} \|k(u - u_{\mathcal{T}})\|^2)
\end{aligned}$$

and the assertion follows.

**Part 2.** We will prove (3.4). Integration by parts leads to

$$\begin{aligned}
\tilde{a}_{\mathcal{T}}(v, w) &= (\nabla_{\mathcal{T}} v, \nabla_{\mathcal{T}} w) + (k^2 v, w) + \mathfrak{i}(kv, w)_{\partial\Omega} \\
&= ((-\Delta_{\mathcal{T}} + k^2)v, w) + \sum_{K \in \mathcal{T}} (\partial_{\mathbf{n}_K} v, w)_{\partial K} + \mathfrak{i}(kv, w)_{\partial\Omega} \\
&= ((-\Delta_{\mathcal{T}} - k^2)v, w) + 2(k^2 v, w) + ((\partial_{\mathbf{n}} + \mathfrak{i}k)v, w)_{\partial\Omega} \\
&\quad + (\llbracket \nabla_{\mathcal{T}} v \rrbracket_N, \{w\})_{\mathfrak{S}^I} + (\{\nabla_{\mathcal{T}} v\}, \llbracket w \rrbracket_N)_{\mathfrak{S}^I}.
\end{aligned}$$

Since  $u$  is a solution of (2.2) it holds

$$(-\Delta_{\mathcal{T}} - k^2)(u - u_{\mathcal{T}}) = (\Delta_{\mathcal{T}} + k^2)u_{\mathcal{T}} + f \quad \text{and} \quad (\partial_{\mathbf{n}} + \mathfrak{i}k)(u - u_{\mathcal{T}}) = g - (\partial_{\mathbf{n}} + \mathfrak{i}k)u_{\mathcal{T}}.$$

For test functions  $\varphi \in H^1(\Omega)$  we have  $\llbracket \varphi \rrbracket = 0$ ,  $\{\varphi\} = \varphi$  and  $u \in H^{3/2+\varepsilon}(\Omega)$  implies  $\llbracket u \rrbracket = \llbracket \nabla_{\mathcal{T}} u \rrbracket = 0$  on interior edges. Therefore

$$\begin{aligned}
\tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi) &= ((\Delta_{\mathcal{T}} + k^2)u_{\mathcal{T}} + f, \varphi) - (\llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N, \varphi)_{\mathfrak{S}^I} \\
&\quad + (g - (\partial_{\mathbf{n}} + \mathfrak{i}k)u_{\mathcal{T}}, \varphi)_{\partial\Omega} + 2(k^2(u - u_{\mathcal{T}}), \varphi).
\end{aligned} \tag{3.8}$$

We choose  $u_{\mathcal{T}}^* \in S_{\mathcal{T}}^{\mathfrak{p}} \cap C^1(\Omega)$  as the conforming approximant of  $u_{\mathcal{T}}$  as in Corollary A.4 to obtain

$$\|u - u_{\mathcal{T}}\|_{\tilde{a}} \leq \|u - u_{\mathcal{T}}^*\|_{\tilde{a}} + \frac{C}{\mathfrak{a}} C_{\text{conf}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}. \tag{3.9}$$

To estimate the first term in (3.9) we define the set

$$\Phi := \left\{ \varphi \in H^1(\Omega) \cap H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega) : \|\varphi\|_{\tilde{a}} \leq 1 \right\}.$$

Let  $I_1^{\text{hp}} : H^1(\Omega) \rightarrow S_{\mathcal{T}}^{\mathfrak{p}} \cap C^1(\Omega)$  be the interpolation operator as in Theorem A.2. Then,

$(u - u_{\mathcal{T}}^*) / \|u - u_{\mathcal{T}}^*\|_{\tilde{a}} \in \Phi$  and we obtain again with Corollary A.4

$$\begin{aligned}
\|u - u_{\mathcal{T}}^*\|_{\tilde{a}} &\leq \sup_{\varphi \in \Phi} |\tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}^*, \varphi)| \\
&\leq \sup_{\varphi \in \Phi} |\tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi)| + \sup_{\varphi \in \Phi} |\tilde{a}_{\mathcal{T}}(u_{\mathcal{T}} - u_{\mathcal{T}}^*, \varphi)| \\
&\leq \sup_{\varphi \in \Phi} |\tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi)| + \sup_{\varphi \in \Phi} \|u_{\mathcal{T}} - u_{\mathcal{T}}^*\|_{\tilde{a}} \|\varphi\|_{\tilde{a}} \\
&\leq \sup_{\varphi \in \Phi} \left| \tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi) - \underbrace{a_{\mathcal{T}}(u - u_{\mathcal{T}}, I_1^{\text{hp}} \varphi)}_{=0} \right| + \frac{C}{\mathfrak{a}} C_{\text{conf}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}. \quad (3.10)
\end{aligned}$$

Next, we use the representations (3.8) of  $\tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi)$  and (3.3) of  $a_{\mathcal{T}}(u - u_{\mathcal{T}}, I_1^{\text{hp}} \varphi)$  to derive the following expression for the supremum in (3.10)

$$\begin{aligned}
&\tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi) - a_{\mathcal{T}}(u - u_{\mathcal{T}}, I_1^{\text{hp}} \varphi) = ((\Delta_{\mathcal{T}} + k^2) u_{\mathcal{T}} + f, \varphi) - (\llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N, \varphi)_{\mathfrak{S}^I} \\
&\quad + (g - (\partial_{\mathbf{n}} + \mathfrak{i}k) u_{\mathcal{T}}, \varphi)_{\partial\Omega} + 2(k^2(u - u_{\mathcal{T}}), \varphi) \\
&\quad - (f + \Delta_{\mathcal{T}} u_{\mathcal{T}} + k^2 u_{\mathcal{T}}, I_1^{\text{hp}} \varphi) + (\llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N, \{I_1^{\text{hp}} \varphi\})_{\mathfrak{S}^I} - (\llbracket u_{\mathcal{T}} \rrbracket_N, \{\nabla_{\mathcal{T}} I_1^{\text{hp}} \varphi\})_{\mathfrak{S}^I} \\
&\quad - \left( \left(1 - \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}}\right) (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i}k u_{\mathcal{T}}), I_1^{\text{hp}} \varphi \right)_{\partial\Omega} + \left( \frac{\mathfrak{d}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i}k u_{\mathcal{T}}), \partial_{\mathbf{n}} I_1^{\text{hp}} \varphi \right)_{\partial\Omega} \\
&\quad + \left( \mathfrak{i} \mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \llbracket u_{\mathcal{T}} \rrbracket_N, \llbracket I_1^{\text{hp}} \varphi \rrbracket_N \right)_{\mathfrak{S}^I} - \left( \frac{\mathfrak{b}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N, \llbracket \nabla_{\mathcal{T}} I_1^{\text{hp}} \varphi \rrbracket_N \right)_{\mathfrak{S}^I} \\
&= ((\Delta_{\mathcal{T}} + k^2) u_{\mathcal{T}} + f, \varphi - I_1^{\text{hp}} \varphi) + 2(k^2(u - u_{\mathcal{T}}), \varphi) - (\llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N, \varphi - I_1^{\text{hp}} \varphi)_{\mathfrak{S}^I} \\
&\quad + (g - (\partial_{\mathbf{n}} + \mathfrak{i}k) u_{\mathcal{T}}, \varphi - I_1^{\text{hp}} \varphi)_{\partial\Omega} + \left( \mathfrak{d} \frac{k\mathfrak{h}}{\mathfrak{p}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i}k u_{\mathcal{T}}), I_1^{\text{hp}} \varphi \right)_{\partial\Omega} \\
&\quad + \left( \frac{\mathfrak{d}\mathfrak{h}}{\mathfrak{i}\mathfrak{p}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i}k u_{\mathcal{T}}), \partial_{\mathbf{n}} I_1^{\text{hp}} \varphi \right)_{\partial\Omega} - (\llbracket u_{\mathcal{T}} \rrbracket_N, \nabla_{\mathcal{T}} I_1^{\text{hp}} \varphi)_{\mathfrak{S}^I}. \quad (3.11)
\end{aligned}$$

We denote the terms after the equal sign in (3.11) by  $T_1, \dots, T_7$  and separately estimate them in the sequel. The constants  $C$  only depend on  $\mathfrak{b}, \mathfrak{d}$  in (2.8), the shape regularity of the mesh, and the constant  $C$  in (A.1).

@ $T_1$  :

$$\left| ((\Delta_{\mathcal{T}} + k^2) u_{\mathcal{T}} + f, \varphi - I_1^{\text{hp}} \varphi) \right| \stackrel{(\text{A.1a})}{\leq} C \left\| \frac{\mathfrak{h}}{\mathfrak{p}} (\Delta_{\mathcal{T}} u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f) \right\| \|\nabla \varphi\|.$$

@ $T_3$  :

$$\begin{aligned}
\left| (\llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N, \varphi - I_1^{\text{hp}} \varphi)_{\mathfrak{S}^I} \right| &\leq \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left( \sum_{e \in \mathcal{E}^I} \frac{p_e}{\mathfrak{b} h_e} \left\| (\varphi - I_1^{\text{hp}} \varphi) \right\|_e^2 \right)^{1/2} \\
&\stackrel{(\text{A.1b})}{\leq} C \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \|\nabla \varphi\|. \quad (3.12)
\end{aligned}$$

@ $T_7$  : Using  $H^1$ -stability of  $I_1^{\text{hp}}$ , we obtain

$$\begin{aligned}
\left| \left( \llbracket u_{\mathcal{T}} \rrbracket_N, \nabla_{\mathcal{T}} I_1^{\text{hp}} \varphi \right)_{\mathfrak{S}^I} \right| &\leq \frac{1}{\sqrt{\mathfrak{a}}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}} \nabla_{\mathcal{T}} I_1^{\text{hp}} \varphi \right\|_{\mathfrak{S}^I} \\
&\leq C_1 \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \nabla_{\mathcal{T}} I_1^{\text{hp}} \varphi \right\| \\
&\leq C_1 C_2 \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \|\nabla \varphi\|,
\end{aligned} \tag{3.13}$$

where  $C_1$  depends on the constant in an  $hp$ -explicit inverse estimate for polynomials (see [44, Thm. 4.76]).

@ $T_4$ :

$$\left| \left( g - (\partial_{\mathbf{n}} + \mathbf{i} k) u_{\mathcal{T}}, \varphi - I_1^{\text{hp}} \varphi \right)_{\partial\Omega} \right| \stackrel{(\text{A.1b})}{\leq} C \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - (\partial_{\mathbf{n}} + \mathbf{i} k) u_{\mathcal{T}}) \right\|_{\partial\Omega} \|\nabla \varphi\|.$$

@ $T_5$ : We use  $\mathfrak{d} k^{1/2} h_e / p_e \leq \mathfrak{d} M_{\frac{\mathfrak{kh}}{\mathfrak{p}}}^{1/2} (h_e / p_e)^{1/2}$  and obtain

$$\begin{aligned}
\left\| \mathfrak{d} \frac{k \mathfrak{h}}{\mathfrak{p}} I_1^{\text{hp}} \varphi \right\|_e &\leq \left\| \mathfrak{d} \frac{k \mathfrak{h}}{\mathfrak{p}} \varphi \right\|_e + \left\| \mathfrak{d} \frac{k \mathfrak{h}}{\mathfrak{p}} (\varphi - I_1^{\text{hp}} \varphi) \right\|_e \\
&\stackrel{(\text{A.1b})}{\leq} C \mathfrak{d} \left( \frac{h_e}{p_e} \right)^{1/2} \left( M_{\frac{\mathfrak{kh}}{\mathfrak{p}}}^{1/2} \|k^{1/2} \varphi\|_{L^2(e)} + M_{\frac{\mathfrak{kh}}{\mathfrak{p}}} \|\nabla \varphi\|_{L^2(\omega_e)} \right).
\end{aligned}$$

This leads to

$$\begin{aligned}
|T_5| &\leq 2C \mathfrak{d} \left( 1 + M_{\frac{\mathfrak{kh}}{\mathfrak{p}}} \right) \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \right\|_{\partial\Omega} (\|k^{1/2} \varphi\|_{\partial\Omega} + \|\nabla \varphi\|) \\
&\leq \tilde{C} \mathfrak{d} \left( 1 + M_{\frac{\mathfrak{kh}}{\mathfrak{p}}} \right) \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \right\|_{\partial\Omega} \|\varphi\|_{\tilde{a}}.
\end{aligned}$$

@ $T_6$ : We obtain similarly as in (3.13)

$$\begin{aligned}
\left| \left( \frac{\mathfrak{d} \mathfrak{h}}{\mathbf{i} \mathfrak{p}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}), \partial_{\mathbf{n}} I_1^{\text{hp}} \varphi \right)_{\partial\Omega} \right| &\leq C \mathfrak{d} \|g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}\|_{\partial\Omega} \left\| \frac{\mathfrak{h}}{\mathfrak{p}} \partial_{\mathbf{n}} I_1^{\text{hp}} \varphi \right\|_{\partial\Omega} \\
&\leq C \mathfrak{d} \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \right\|_{\partial\Omega} \|\nabla \varphi\|.
\end{aligned}$$

These estimates allow to bound the expression in the supremum of (3.10) by

$$\begin{aligned}
& \left| \tilde{a}_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi) - a_{\mathcal{T}}(u - u_{\mathcal{T}}, I_1^{\text{hp}} \varphi) \right| \leq 2 \|k(u - u_{\mathcal{T}})\| \|k\varphi\| \\
& + C \left( \left\| \frac{\mathfrak{h}}{\mathfrak{p}} (\Delta_{\mathcal{T}} u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f) \right\| \|\nabla \varphi\| + \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} [\nabla_{\mathcal{T}} u_{\mathcal{T}}]_N \right\|_{\mathfrak{S}^I} \|\nabla \varphi\| \right. \\
& + \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - (\partial_{\mathbf{n}} + \mathbf{i} k) u_{\mathcal{T}}) \right\|_{\partial \Omega} \|\nabla \varphi\| \\
& + \left(1 + M_{\frac{\mathbf{k}\mathbf{h}}{\mathfrak{p}}}\right) \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \right\|_{\partial \Omega} \|\varphi\|_{\tilde{a}} \\
& \left. + \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \right\|_{\partial \Omega} \|\nabla \varphi\| + \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} [u_{\mathcal{T}}]_N \right\|_{\mathfrak{S}^I} \|\nabla \varphi\| \right).
\end{aligned}$$

The combination of (3.9), (3.10) with the definitions of  $\eta_R$ ,  $\eta_E$ ,  $\eta_J$  leads to

$$\begin{aligned}
\|u - u_{\mathcal{T}}\|_{\tilde{a}} & \leq 2 \|k(u - u_{\mathcal{T}})\| + C \left( \left\| \frac{\mathfrak{h}}{\mathfrak{p}} (\Delta_{\mathcal{T}} u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f) \right\| \right. \\
& + C_{\text{conf}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} [u_{\mathcal{T}}]_N \right\|_{\mathfrak{S}^I} + \left\| \sqrt{\mathfrak{b} \frac{\mathfrak{h}}{\mathfrak{p}}} [\nabla_{\mathcal{T}} u_{\mathcal{T}}]_N \right\|_{\mathfrak{S}^I} \\
& \left. + \left(1 + M_{\frac{\mathbf{k}\mathbf{h}}{\mathfrak{p}}}\right) \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \right\|_{\partial \Omega} \right) \\
& \leq CC_{\text{conf}} (\eta_R^2 + \eta_E^2 + \eta_J^2)^{1/2} + 2k \|u - u_{\mathcal{T}}\|.
\end{aligned}$$

This concludes the proof for  $p_{\mathcal{T}} \geq 5$ .

For  $1 \leq p_{\mathcal{T}} < 5$  we have to employ  $I_1^{\text{hp},0}$  instead of  $I_1^{\text{hp}}$  (cf. Theorem A.2). For the details of this case we refer to [47, Rem. 4.1.4].  $\square$

To prove the reliability estimate it remains to bound the term  $\|k(u - u_{\mathcal{T}})\|$  by the estimator. We will show that  $\|k(u - u_{\mathcal{T}})\|$  is bounded (modulo constants) by the product of  $\eta(u_{\mathcal{T}})$  with the adjoint approximation property  $\sigma_k^*(S)$  (see (2.12)).

**Lemma 3.4.** *Let the assumptions of Lemma 3.2 be satisfied. There exists a constant  $C$  solely depending on  $\rho_{\mathcal{T}}$ ,  $\mathfrak{b}$ ,  $\mathfrak{d}$ , and  $\Omega$  such that*

$$\|k(u - u_{\mathcal{T}})\| \leq C \eta(u_{\mathcal{T}}) \sigma_k^*(S_{\mathcal{T}}^{\mathfrak{p}})$$

with  $\sigma_k^*(S_{\mathcal{T}}^{\mathfrak{p}})$  as in (2.12).

*Proof. Part 1.* We will prove

$$|a_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi)| \leq C \eta(u_{\mathcal{T}}) \|\varphi\|_{\text{dG}^+} \quad \forall \varphi \in H^1(\Omega) \cap H_{\mathcal{T}}^{3/2+\varepsilon}(\Omega). \quad (3.14)$$

Note that  $\llbracket I_1^{\text{hp}} \varphi \rrbracket_N = \llbracket \nabla I_1^{\text{hp}} \varphi \rrbracket_N = 0$ . We employ Lemma 3.3 and the estimates for  $T_1, \dots, T_7$  in the proof of Lemma 3.2 to obtain

$$\begin{aligned}
|a_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi)| &= \left| a_{\mathcal{T}}(u - u_{\mathcal{T}}, \varphi - I_1^{\text{hp}} \varphi) \right| \tag{3.15} \\
&\leq C \left( \left\| \frac{\mathfrak{h}}{\mathfrak{p}} (\Delta_{\mathcal{T}} u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f) \right\| \|\nabla \varphi\| + \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \sqrt{\frac{\mathfrak{p}}{\mathfrak{b}\mathfrak{h}}} (\varphi - I_1^{\text{hp}} \varphi) \right\|_{\mathfrak{S}^I} \right. \\
&\quad + \left\| \sqrt{\frac{\mathfrak{p}^2}{\mathfrak{a}\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \left( \frac{\mathfrak{p}^2}{\mathfrak{a}\mathfrak{h}} \right)^{-1/2} \left\{ \nabla (\varphi - I_1^{\text{hp}} \varphi) \right\} \right\|_{\mathfrak{S}^I} \\
&\quad + \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i} k u_{\mathcal{T}}) \right\|_{\partial\Omega} \left\| \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}} \partial_{\mathbf{n}} (\varphi - I_1^{\text{hp}} \varphi) \right\|_{\partial\Omega} \\
&\quad \left. + \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i} k u_{\mathcal{T}}) \right\|_{\partial\Omega} \|\nabla \varphi\| + \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} \varphi \rrbracket_N \right\|_{\mathfrak{S}^I} \right).
\end{aligned}$$

Note that

$$\begin{aligned}
\left\| \left( \frac{\mathfrak{p}^2}{\mathfrak{a}\mathfrak{h}} \right)^{-1/2} \left\{ \nabla (\varphi - I_1^{\text{hp}} \varphi) \right\} \right\|_{\mathfrak{S}^I} &\leq C \left( \left\| \left( \frac{\mathfrak{p}^2}{\mathfrak{a}\mathfrak{h}} \right)^{-1/2} \left\{ \nabla \varphi \right\} \right\|_{\mathfrak{S}^I} + \left\| \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}\sqrt{\mathfrak{a}}} \nabla I_1^{\text{hp}} \varphi \right\|_{\mathfrak{S}^I} \right) \tag{3.16} \\
&\stackrel{(3.13)}{\leq} C \left( \left\| \left( \frac{\mathfrak{p}^2}{\mathfrak{a}\mathfrak{h}} \right)^{-1/2} \left\{ \nabla \varphi \right\} \right\|_{\mathfrak{S}^I} + \|\nabla \varphi\| \right) \leq C \|\varphi\|_{\text{dG}^+}.
\end{aligned}$$

We also use

$$\begin{aligned}
\left\| \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}} \partial_{\mathbf{n}} (\varphi - I_1^{\text{hp}} \varphi) \right\|_{\partial\Omega} &\leq C \|\varphi\|_{\text{dG}} + \left\| \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}} \partial_{\mathbf{n}} I_1^{\text{hp}} \varphi \right\|_{\partial\Omega} \tag{3.17} \\
&\stackrel{(3.13)}{\leq} C (\|\varphi\|_{\text{dG}} + \|\nabla \varphi\|) \leq C \|\varphi\|_{\text{dG}}.
\end{aligned}$$

From the combination of (3.12), (3.15), (3.16), (3.17) with the definition of the error estimator we conclude that (3.14) holds.

**Part 2.** We will derive the assertion by using (3.14) and an Aubin-Nitsche argument. For  $Q_k^*$  as defined after (2.11), let  $z := Q_k^*(k^2(u - u_{\mathcal{T}}))$ . Furthermore let  $z_S \in S_{\mathcal{T}}^{\mathfrak{p}}$  be the best approximation of  $z$  in the finite element space with respect to the norm  $\|\cdot\|_{\text{dG}^+}$ , i.e.

$$\|z - z_S\|_{\text{dG}^+} = \inf_{w \in S_{\mathcal{T}}^{\mathfrak{p}}} \|z - w\|_{\text{dG}^+}.$$

With Lemma 2.6 it follows

$$\|k(u - u_{\mathcal{T}})\|^2 = (u - u_{\mathcal{T}}, k^2(u - u_{\mathcal{T}})) = a_{\mathcal{T}}(u - u_{\mathcal{T}}, z) = a_{\mathcal{T}}(u - u_{\mathcal{T}}, z - z_S).$$

By using the adjoint approximation property (2.12) we get

$$\|z - z_S\|_{\text{dG}^+} = \inf_{w \in S_{\mathcal{T}}^{\mathfrak{p}}} \|Q_k^*(k^2(u - u_{\mathcal{T}})) - w\|_{\text{dG}^+} \leq \sigma_k^*(S_{\mathcal{T}}^{\mathfrak{p}}) \|k(u - u_{\mathcal{T}})\|.$$



Employing (3.14) we end up with

$$\begin{aligned}\|k(u - u_{\mathcal{T}})\|^2 &= a_{\mathcal{T}}(u - u_{\mathcal{T}}, z - z_S) \leq C\eta(u_{\mathcal{T}}) \|z - z_S\|_{\text{dG}^+} \\ &\leq C\eta(u_{\mathcal{T}}) \sigma_k^*(S_{\mathcal{T}}^{\mathbf{p}}) \|k(u - u_{\mathcal{T}})\|,\end{aligned}$$

which implies the assertion.  $\square$

The next theorem states the reliability estimate for our a posteriori error estimator which is explicit in the discretization parameters  $h$ ,  $p$ , and the wavenumber  $k$ . Its proof is a simple combination of Lemma 3.2 and Lemma 3.4. For later use we define a modified error estimator where  $f$  and  $g$  are replaced by projections to polynomial spaces and *data oscillations*. In order to obtain reliability *and* efficiency for the *same* error estimator (up to data oscillations) we will also state reliability for the modified error estimator in the following theorem; the latter follows from the reliability of the original error estimator  $\eta$  (cf. [47, Thm. 4.1.10]) via a triangle inequality.

**Definition 3.5.** For  $f \in L^2(\Omega)$ , let  $f_{\mathcal{T}}$  be the simplex-wise polynomial function with  $f_{\mathcal{T}}|_K$  denoting the  $L^2(K)$  orthogonal projection of  $f|_K$  onto  $\mathbb{P}_{p_K}(K)$ . For  $g \in L^2(\partial^B K)$ , let  $g_{\partial^B K} \in L^2(\partial^B K)$  be the edge-wise polynomial function with  $g_{\partial^B K}|_e$  denoting the  $L^2(e)$  orthogonal projection of  $g|_e$  onto  $\mathbb{P}_{p_K}(e)$ . The data oscillations are given for  $K \in \mathcal{T}$  by

$$\text{osc}_K := \left( \left\| \frac{h_K}{p_K} (f - f_{\mathcal{T}}) \right\|_{L^2(K)}^2 + \left\| \sqrt{\mathbf{h}} (g - g_{\partial^B K}) \right\|_{\partial^B K}^2 \right)^{1/2}$$

and

$$\text{osc}_{\mathcal{T}} := \left( \sum_{K \in \mathcal{T}} \text{osc}_K^2 \right)^{1/2}.$$

The local error estimators  $\tilde{\eta}_K$ ,  $\tilde{\eta}_{R_K}$ ,  $\tilde{\eta}_{E_K}$  are given by replacing  $f$  by  $f_{\mathcal{T}}$  in (3.1b),  $g$  by  $g_{\partial^B K}$  in (3.1b), and  $\eta_{R_K}$  and  $\eta_{E_K}$  by  $\tilde{\eta}_{R_K}$  and  $\tilde{\eta}_{E_K}$  in (3.1a). The global estimators  $\tilde{\eta}_R$ ,  $\tilde{\eta}_E$ , and  $\tilde{\eta}$  are given by replacing  $\eta_{R_K}$  and  $\eta_{E_K}$  by  $\tilde{\eta}_{R_K}$  and  $\tilde{\eta}_{E_K}$  in (3.2b) and  $\eta_R$  and  $\eta_E$  by  $\tilde{\eta}_R$  and  $\tilde{\eta}_E$  in (3.2a).

**Theorem 3.6.** Let  $\mathcal{T}$  be a shape regular, conforming simplicial finite element mesh of the polygonal Lipschitz domain  $\Omega \subseteq \mathbb{R}^2$  and let the polynomial degree function  $\mathbf{p}$  satisfies (2.4) and  $p_{\mathcal{T}} \geq 1$ . Assume that  $k > 1$  is constant. Let  $u \in H^{3/2+\varepsilon}(\Omega)$  be the solution of (2.2) for some  $\varepsilon > 0$  and assume that  $u_{\mathcal{T}} \in S_{\mathcal{T}}^{\mathbf{p}}$  solves (2.8) with  $\mathbf{a} \geq 1$ . Then, there exists a constant  $C > 0$  solely depending on  $\rho_{\mathcal{T}}$ ,  $\mathbf{b}$ ,  $\mathbf{d}$ , and  $\Omega$  such that

$$\|u - u_{\mathcal{T}}\|_{\text{dG}} \leq C \sqrt{1 + M_{\frac{\mathbf{kh}}{\mathbf{p}}}} \left( 1 + M_{\frac{\mathbf{kh}}{\mathbf{p}}} + \sigma_k^*(S_{\mathcal{T}}^{\mathbf{p}}) \right) \eta(u_{\mathcal{T}}).$$

For the modified error estimator it holds

$$\|u - u_{\mathcal{T}}\|_{\text{dG}} \leq C \left( 1 + M_{\frac{\mathbf{kh}}{\mathbf{p}}} \right)^{3/2} \left( 1 + \sigma_k^*(S_{\mathcal{T}}^{\mathbf{p}}) \right) (\tilde{\eta}(u_{\mathcal{T}}) + \text{osc}_{\mathcal{T}}(u_{\mathcal{T}})).$$

### 3.3 Efficiency

The reliability estimate in the form of Theorem 3.6 shows that the error estimator (modulo a constant  $C$  which only depends on  $\rho_{\mathcal{T}}$ ,  $\mathbf{b}$ ,  $\mathbf{d}$ , and  $\Omega$ ) controls the error of the dG-approximation  $u_{\mathcal{T}}$  in a reliable way. This estimate can be used as a stopping criterion within an adaptive discretization process.

In this section we are concerned with the *efficiency* of the error estimator which ensures that the error estimator converges with the same rate as the true error. Efficiency can be proved locally, i.e., the localized error estimator is estimated by the localized error. For the proof, we employ ideas which have been developed for conforming finite element methods in [39] and for dG-methods, e.g., in [27, Thm. 3.2]. As is common for efficiency estimates one has to deal with *data oscillations*.

**Theorem 3.7.** *Let the assumptions of Theorem 3.6 be satisfied. There exists a constant independent of  $k$ ,  $h_K$ ,  $p_K$  such that the modified local internal residual can be estimated by*

$$\tilde{\eta}_{R_K} \leq Cp_K \left( \|\nabla(u - u_{\mathcal{T}})\|_{L^2(K)} + M_{\frac{\mathbf{kh}}{\mathbf{p}}} \|k(u - u_{\mathcal{T}})\|_{L^2(K)} + \left\| \frac{h_K}{p_K} (f - f_{\mathcal{T}}) \right\|_{L^2(K)} \right). \quad (3.18a)$$

For the gradient jumps in the error estimator it holds

$$\begin{aligned} \left\| \sqrt{\mathbf{b} \frac{\mathbf{h}}{\mathbf{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_e &\leq Cp_e^{3/2} \left( \|\nabla(u - u_{\mathcal{T}})\|_{L^2(\omega_e)} + M_{\frac{\mathbf{kh}}{\mathbf{p}}} \|k(u - u_{\mathcal{T}})\|_{L^2(\omega_e)} \right. \\ &\quad \left. + \left\| \frac{h_e}{p_e} (f - f_{p_e}) \right\|_{L^2(\omega_e)} \right). \end{aligned} \quad (3.18b)$$

For the modified local edge residuals it holds

$$\begin{aligned} \tilde{\eta}_{E_K} &\leq Cp_K^2 \left( \|\nabla(u - u_{\mathcal{T}})\|_{L^2(\omega_K)} + M_{\frac{\mathbf{kh}}{\mathbf{p}}} \|k(u - u_{\mathcal{T}})\|_{L^2(\omega_K)} + \left\| \frac{h_K}{p_K} (f - f_{\mathcal{T}}) \right\|_{L^2(\omega_K)} \right. \\ &\quad \left. + \sqrt{M_{\frac{\mathbf{kh}}{\mathbf{p}}}} \left\| \sqrt{\frac{k}{p_K}} (u - u_{\mathcal{T}}) \right\|_{\partial^B K} + \left\| \frac{h_K^{1/2}}{p_K} (g - g_{\partial^B K}) \right\|_{\partial^B K} \right). \end{aligned} \quad (3.18c)$$

Let  $\mathbf{a} \geq C_{\mathbf{a}} > 0$  for some sufficiently large constant  $C_{\mathbf{a}}$  depending only on the shape regularity of the mesh. Then, there exists a constant  $C > 0$  such that

$$\begin{aligned} \eta_J^2 &\leq C \sum_{K \in \mathcal{T}} p_K^4 \left( M_{\frac{\mathbf{kh}}{\mathbf{p}}}^2 \|k(u - u_{\mathcal{T}})\|_{L^2(K)}^2 + \left\| \frac{\mathbf{h}}{\mathbf{p}} (f - f_{\mathcal{T}}) \right\|_{L^2(K)}^2 + \|\nabla(u - u_{\mathcal{T}})\|_{L^2(K)}^2 \right. \\ &\quad \left. + M_{\frac{\mathbf{kh}}{\mathbf{p}}} \left\| \sqrt{\frac{k}{p_K}} (u - u_{\mathcal{T}}) \right\|_{\partial^B K}^2 + \left\| \frac{\sqrt{h_K}}{p_K} (g - g_{\partial^B K}) \right\|_{\partial^B K}^2 \right). \end{aligned} \quad (3.18d)$$

*Proof.* The proof of these estimates follow the ideas of [39] (see also [16, Proof of Thm. 4.12]) and are worked out in detail in [47, Sec. 4.2]. Here we prove exemplarily (3.18c) and (3.18d).

**Proof of (3.18c).**

We consider the estimate for the edge residuals and start by introducing an *edge bubble function*. We define  $\hat{e} := [0, 1]$  and  $\Phi_{\hat{e}} : [0, 1] \rightarrow \mathbb{R}$  by  $\Phi_{\hat{e}}(x) := x(1-x)$ . For  $K \in \mathcal{T}$ , let  $F_K : \hat{K} \rightarrow K$  be a usual affine pullback to the reference element  $\hat{K} := \text{conv} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$ . For  $e \in \mathcal{E}(K)$ , we may choose  $F_K$  in such a way that  $F_e := F_K|_{\hat{e}} : \hat{e} \rightarrow e$ . Then we define  $\Phi_e : e \rightarrow \mathbb{R}$  and the global version  $\Phi_{\mathcal{E}} : \mathfrak{S} \rightarrow \mathbb{R}$  by

$$\Phi_e := c_e \Phi_{\hat{e}} \circ F_e^{-1} \quad \text{with } c_e \in \mathbb{R} \text{ such that } \int_e \Phi_e = h_e \quad \text{and} \quad \forall e \in \mathcal{E} : \quad \Phi_{\mathcal{E}}|_e := \Phi_e.$$

For  $\zeta \in [0, 1]$ , we introduce

$$\begin{aligned} \tilde{\eta}_{\zeta; E_K}(u_{\mathcal{T}}) := & \left( \frac{1}{2} \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \Phi_{\mathcal{E}}^{\zeta/2} \right\|_{\partial^I K}^2 \right. \\ & \left. + \left\| \sqrt{\mathfrak{h}} (g_{\partial^B K} - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}) \Phi_{\mathcal{E}}^{\zeta/2} \right\|_{\partial^B K}^2 \right)^{1/2} \end{aligned} \quad (3.19)$$

and note that  $\tilde{\eta}_{0, E_K}(u_{\mathcal{T}}) = \tilde{\eta}_{E_K}(u_{\mathcal{T}})$ .

For the remaining part of the proof we follow the arguments in [39, Lem. 3.5] and consider first the second term in the right-hand side of (3.19). Let first  $\zeta \in ]\frac{1}{2}, 1]$ . To estimate the second term we employ a certain extension of  $\Phi_{\hat{e}}^{\zeta}$  to  $K$  whose existence is proved in [39, Lem. 2.6] and is stated as follows: Let  $\hat{K}$  be the reference element and let  $\hat{e} = [0, 1] \times \{0\}$ . Let  $\zeta \in ]\frac{1}{2}, 1]$ . Then there exists  $C = C(\zeta) > 0$  such that, for any  $\varepsilon \in ]0, 1]$ ,  $p \in \mathbb{N}$ , and  $\hat{q} \in \mathbb{P}_p(\hat{e})$ , there exists an extension  $v_{\hat{e}} \in H^1(\hat{K})$  of  $\hat{q} \Phi_{\hat{e}}^{\zeta}$  with

$$v_{\hat{e}}|_{\hat{e}} = \hat{q} \Phi_{\hat{e}}^{\zeta} \quad \text{and} \quad v_{\hat{e}}|_{\partial \hat{K} \setminus \hat{e}} = 0, \quad (3.20a)$$

$$\|v_{\hat{e}}\|_{L^2(\hat{K})}^2 \leq C \varepsilon \left\| \hat{q} \Phi_{\hat{e}}^{\zeta/2} \right\|_{\hat{e}}^2, \quad (3.20b)$$

$$\|\nabla v_{\hat{e}}\|_{L^2(\hat{K})}^2 \leq C (\varepsilon p^{2(2-\zeta)} + \varepsilon^{-1}) \left\| \hat{q} \Phi_{\hat{e}}^{\zeta/2} \right\|_{\hat{e}}^2. \quad (3.20c)$$

For  $e \subset \partial^B K$ , choose the affine pullback  $F_K$  such that, for  $F_e := F_K|_e$ , it holds  $F_e(\hat{e}) = e$ . We set  $q := g_{\partial^B K} - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}$ , denote the pullback by  $\hat{q} := q \circ F_e \in \mathbb{P}_{p_K}$ , and let  $v_{\hat{e}}$  denote the above extension for this choice of  $\hat{q}$ . Then  $w_e := v_{\hat{e}} \circ F_K \in H^1(K)$  and satisfies  $w_e|_{\partial K \setminus e} = 0$ . Thus, we obtain with  $\partial_{\mathbf{n}} u + \mathbf{i} k u = g$  on  $\partial \Omega$

$$\begin{aligned} \|q \Phi_e^{\zeta/2}\|_e^2 &= (g_{\partial^B K} - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathbf{i} k u_{\mathcal{T}}, w_e)_e \\ &= (\partial_{\mathbf{n}}(u - u_{\mathcal{T}}), w_e)_e + (\mathbf{i} k(u - u_{\mathcal{T}}), w_e)_e + (g_{\partial^B K} - g, w_e)_e. \end{aligned} \quad (3.21)$$

We estimate these terms separately and start with the last one and obtain by using that  $\Phi_{\mathcal{E}}^{\zeta/2}$  is bounded pointwise by a constant  $C > 0$  uniformly in  $\zeta \in [0, 1]$  and  $x \in \mathfrak{S}$

$$\begin{aligned} (g_{\partial^B K} - g, w_e)_e &\leq \|g_{\partial^B K} - g\|_e \|w\|_e = \|g_{\partial^B K} - g\|_e \|q \Phi_e^{\zeta/2}\|_e \\ &\leq C \|g_{\partial^B K} - g\|_e \|q \Phi_e^{\zeta/2}\|_e. \end{aligned}$$

For the second term of the right-hand side in (3.21) we derive in a similar fashion

$$(\mathbf{i} k(u - u_{\mathcal{T}}), w_e)_e \leq C \|k(u - u_{\mathcal{T}})\|_e \|q\Phi_e^{\zeta/2}\|_e.$$

For the first term in (3.21) we get

$$\begin{aligned} (\partial_{\mathbf{n}}(u - u_{\mathcal{T}}), w_e)_e &= (\partial_{\mathbf{n}}(u - u_{\mathcal{T}}), w_e)_{\partial K} \\ &= (\nabla(u - u_{\mathcal{T}}), \nabla w_e)_{L^2(K)} + (\Delta(u - u_{\mathcal{T}}), w_e)_{L^2(K)} \\ &= (\nabla(u - u_{\mathcal{T}}), \nabla w_e)_{L^2(K)} + (k^2(u_{\mathcal{T}} - u), w_e)_{L^2(K)} - (\Delta u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f, w_e)_{L^2(K)} \\ &\leq \|\nabla(u - u_{\mathcal{T}})\|_{L^2(K)} \|\nabla w_e\|_{L^2(K)} + \left( \|k^2(u - u_{\mathcal{T}})\|_{L^2(K)} \right. \\ &\quad \left. + \|\Delta u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f_{\mathcal{T}}\|_{L^2(K)} + \|f - f_{\mathcal{T}}\|_{L^2(K)} \right) \|w_e\|_{L^2(K)}. \end{aligned}$$

By scaling (3.20b), (3.20c) to the triangle  $K$  and estimating  $\|\Delta u_{\mathcal{T}} + k^2 u_{\mathcal{T}} + f_{\mathcal{T}}\|_{L^2(K)} = \frac{p_K}{h_K} \tilde{\eta}_{R_K}(u_{\mathcal{T}})$  via (3.18a), we get

$$\begin{aligned} (\partial_{\mathbf{n}}(u - u_{\mathcal{T}}), w_e)_e &\leq C \|q\Phi_e^{\zeta/2}\|_e \left\{ \left( \frac{\varepsilon p_K^{2(2-\zeta)} + \varepsilon^{-1}}{h_K} \right)^{1/2} \|\nabla(u - u_{\mathcal{T}})\|_{L^2(K)} \right. \\ &\quad \left. + \sqrt{\varepsilon h_K} \left( \|k^2(u - u_{\mathcal{T}})\|_{L^2(K)} + \frac{p_K}{h_K} \tilde{\eta}_{R_K}(u_{\mathcal{T}}) + \|f - f_{\mathcal{T}}\|_{L^2(K)} \right) \right\} \\ &\leq C \|q\Phi_e^{\zeta/2}\|_e \left\{ \left( \frac{\varepsilon p_K^{2(2-\zeta)} + \varepsilon^{-1} + \varepsilon p_K^4}{h_K} \right)^{1/2} \|\nabla(u - u_{\mathcal{T}})\|_{L^2(K)} \right. \\ &\quad \left. + \sqrt{\varepsilon h_K} \left( \|k^2(u - u_{\mathcal{T}})\|_{L^2(K)} + \frac{p_K^2}{h_K} M_{\frac{\mathbf{k}\mathbf{h}}{\mathbf{p}}} \|k(u - u_{\mathcal{T}})\|_{L^2(K)} \right. \right. \\ &\quad \left. \left. + p_K \|f - f_{\mathcal{T}}\|_{L^2(K)} \right) \right\}. \end{aligned}$$

Altogether we have proved (for the choice  $\varepsilon = p_K^{-2}$ )

$$\begin{aligned} \left\| \sqrt{\mathbf{h}} q\Phi_e^{\zeta/2} \right\|_{\partial^B K}^2 &\leq C p_K^2 \left( \|\nabla(u - u_{\mathcal{T}})\|_{L^2(K)}^2 + M_{\frac{\mathbf{k}\mathbf{h}}{\mathbf{p}}}^2 \|k(u - u_{\mathcal{T}})\|_{L^2(K)}^2 \right. \\ &\quad \left. + \left\| \frac{h_K}{p_K} f - f_{\mathcal{T}} \right\|_{L^2(K)}^2 + M_{\frac{\mathbf{k}\mathbf{h}}{\mathbf{p}}}^2 \left\| \sqrt{\frac{k}{p_K}} (u - u_{\mathcal{T}}) \right\|_{\partial^B K}^2 + \left\| \frac{\sqrt{\mathbf{h}}}{p_K} (g_{\partial^B K} - g) \right\|_{\partial^B K}^2 \right). \end{aligned} \quad (3.22)$$

For  $\zeta \in [0, 1/2]$  we obtain from [39, Lem. 2.4 with  $\beta = 1$  and  $\alpha = \zeta$ ]

$$\|q\Phi_e^{\zeta/2}\|_e \leq C p_K^{1-\zeta} \|q\Phi_e^{1/2}\|_e. \quad (3.23)$$

By choosing  $\zeta = 0$  in (3.23) and  $\zeta = 1$  in (3.22) we get

$$\begin{aligned} \left\| \sqrt{\mathfrak{h}} q \right\|_{\partial^B K} &\leq C p_K \left\| q \Phi_{\mathcal{E}}^{1/2} \right\|_{\partial^B K} \leq C p_K^2 \left( \left\| \nabla (u - u_{\mathcal{T}}) \right\|_{L^2(K)} \right. \\ &\quad + M_{\frac{\mathfrak{kh}}{\mathfrak{p}}} \left\| k (u - u_{\mathcal{T}}) \right\|_{L^2(K)} + \left\| \frac{h_K}{p_K} f - f_{\mathcal{T}} \right\|_{L^2(K)} \\ &\quad \left. + \sqrt{M_{\frac{\mathfrak{kh}}{\mathfrak{p}}}} \left\| \sqrt{\frac{k}{p_K}} (u - u_{\mathcal{T}}) \right\|_{\partial^B K} + \left\| \frac{\sqrt{\mathfrak{h}}}{p_K} (g_{\partial^B K} - g) \right\|_{\partial^B K} \right). \end{aligned} \quad (3.24)$$

This finishes the estimate of the second term in the right-hand side of (3.19). The first term can be estimated via (3.18b) and leads to (3.18c).

**Proof of (3.18d).**

**Part 1.** We prove

$$\begin{aligned} \eta_J^2 &\leq C \left( \text{osc}_{\mathcal{T}}^2 + \tilde{\eta}_R^2 + \sum_{K \in \mathcal{T}} \left( \frac{p_K}{2} \left\| \sqrt{\frac{\mathfrak{h}}{\mathfrak{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\partial^I K}^2 \right. \right. \\ &\quad \left. \left. + \left\| \sqrt{\mathfrak{h}} (g_{\partial^B K} - (\partial_{\mathbf{n}} + \mathfrak{i} k) u_{\mathcal{T}}) \right\|_{\partial^B K}^2 \right) \right). \end{aligned} \quad (3.25)$$

Let  $u_{\mathcal{T}}^* \in S_{\mathcal{T}}^{\mathfrak{p}}$  denote the conforming approximant of  $u_{\mathcal{T}}$  (cf. Corollary A.4). Due to Galerkin orthogonality it holds

$$a_{\mathcal{T}}(u - u_{\mathcal{T}}, u_{\mathcal{T}} - u_{\mathcal{T}}^*) = 0. \quad (3.26)$$

The continuity of  $u_{\mathcal{T}}^*$  implies

$$\begin{aligned} \sum_{K \in \mathcal{T}} \eta_{J_K}^2 &= \sum_{K \in \mathcal{T}} \frac{1}{2} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\partial^I K}^2 = \left| \left( \mathfrak{i} \mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \llbracket u_{\mathcal{T}} \rrbracket_N, \llbracket u_{\mathcal{T}} \rrbracket_N \right)_{\mathfrak{S}^I} \right| \\ &= \left| \left( \mathfrak{i} \mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \llbracket u_{\mathcal{T}} \rrbracket_N, \llbracket u_{\mathcal{T}} - u_{\mathcal{T}}^* \rrbracket_N \right)_{\mathfrak{S}^I} \right| \end{aligned}$$

and we combine (3.26) with the representation as in Lemma 3.3 to obtain

$$\begin{aligned} \sum_{K \in \mathcal{T}} \eta_{J_K}^2 &\leq \eta_R \left\| \frac{\mathfrak{p}}{\mathfrak{h}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \right\| + \left\| \sqrt{\mathfrak{h} \mathfrak{d}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| (\mathfrak{h} \mathfrak{d})^{-1/2} \{u_{\mathcal{T}} - u_{\mathcal{T}}^*\} \right\|_{\mathfrak{S}^I} \\ &\quad + \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \left( \mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}} \right)^{-1/2} \{ \nabla_{\mathcal{T}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \} \right\|_{\mathfrak{S}^I} \\ &\quad + \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i} k u_{\mathcal{T}}) \right\|_{\partial \Omega} \left\| \mathfrak{h}^{-1/2} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \right\|_{\partial \Omega} \\ &\quad + \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i} k u_{\mathcal{T}}) \right\|_{\partial \Omega} \left\| \mathfrak{d} \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}} \partial_{\mathbf{n}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \right\|_{\partial \Omega} \\ &\quad + \left\| \sqrt{\mathfrak{b} \mathfrak{h}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \left\| \frac{\sqrt{\mathfrak{b} \mathfrak{h}}}{\mathfrak{p}} \llbracket \nabla_{\mathcal{T}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \rrbracket_N \right\|_{\mathfrak{S}^I}. \end{aligned}$$

The factors which contain  $u_{\mathcal{T}} - u_{\mathcal{T}}^*$  can be estimated by using Theorem A.3 and polynomial inverse estimates

$$\begin{aligned}
\left\| \frac{\mathfrak{p}}{\mathfrak{h}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \right\| &\leq \frac{C}{\sqrt{\mathfrak{a}}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}, \\
\left\| \frac{\{\nabla_{\mathcal{T}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*)\}}{\sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}}} \right\|_{\mathfrak{S}^I} &\leq \frac{C}{\sqrt{\mathfrak{a}}} \|\nabla_{\mathcal{T}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*)\|_{\mathfrak{S}^I} \leq \frac{C}{\mathfrak{a}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}, \\
\left\| \frac{\{u_{\mathcal{T}} - u_{\mathcal{T}}^*\}}{\sqrt{\mathfrak{h}\mathfrak{d}}} \right\|_{\mathfrak{S}^I} + \left\| \frac{u_{\mathcal{T}} - u_{\mathcal{T}}^*}{\sqrt{\mathfrak{h}}} \right\|_{\partial\Omega} &\leq C \left\| \frac{\mathfrak{p}}{\mathfrak{h}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \right\| \leq \frac{C}{\sqrt{\mathfrak{a}}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}, \\
\left\| \mathfrak{d} \frac{\sqrt{\mathfrak{h}}}{\mathfrak{p}} \partial_{\mathbf{n}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \right\|_{\partial\Omega} + \left\| \frac{\sqrt{\mathfrak{b}\mathfrak{h}}}{\mathfrak{p}} \llbracket \nabla_{\mathcal{T}} (u_{\mathcal{T}} - u_{\mathcal{T}}^*) \rrbracket_N \right\|_{\mathfrak{S}^I} \\
&\leq C \|\nabla (u_{\mathcal{T}} - u_{\mathcal{T}}^*)\| \leq \frac{C}{\sqrt{\mathfrak{a}}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}.
\end{aligned}$$

This finally leads to

$$\begin{aligned}
\sum_{K \in \mathcal{T}} \eta_{J_K}^2 &\leq \frac{C}{\sqrt{\mathfrak{a}}} \left( \eta_R + \left\| \sqrt{\mathfrak{h}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} + \left\| \sqrt{\mathfrak{h}} (g - \partial_{\mathbf{n}} u_{\mathcal{T}} - \mathfrak{i} k u_{\mathcal{T}}) \right\|_{\partial\Omega} \right. \\
&\quad \left. + \frac{1}{\sqrt{\mathfrak{a}}} \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket_N \right\|_{\mathfrak{S}^I} \right) \left\| \sqrt{\mathfrak{a} \frac{\mathfrak{p}^2}{\mathfrak{h}}} \llbracket u_{\mathcal{T}} \rrbracket \right\|_{\mathfrak{S}^I}.
\end{aligned}$$

We divide this inequality by the last factor, absorb the last summand in the left-hand side for sufficiently large  $\mathfrak{a}$ , and estimate  $\eta_R \leq \tilde{\eta}_R + \text{osc}_{\mathcal{T}}$ . Thus, we have proved (3.25).

**Part 2.** From (3.25) we will derive (3.18d).

The second term in the right-hand side of (3.25) can be estimated by using (3.18a) while the estimate for the last sum in (3.25) follows from (3.18b) and (3.24).  $\square$

### Remark 3.8.

- a. As is well-known for residual a posteriori error estimation in the context of hp-finite elements, the reliability estimate is robust with respect to the polynomial degree while the efficiency estimate is polluted by powers of  $p_K$  due to inverse inequalities. The theory of [39] allows to shift powers of  $p_K$  in the efficiency estimate to powers of  $p_K$  in the reliability estimate by employing certain powers of bubble functions in the definition of the error estimator. This can also be done for the dG-formulation of the Helmholtz problem and is worked out in [47].
- b. A difference to standard elliptic problems is the appearance of the adjoint approximation property  $\sigma_k^*(S_{\mathcal{T}}^{\mathfrak{p}})$  (cf. (2.13)) in the reliability estimate, and powers of the quantity

$M_{\frac{kh}{p}}$  in (2.6) in both, the efficiency and the reliability estimates. For convex polygonal domains, it can be shown that  $p \geq C_0 \log(k)$  and the resolution condition

$$\frac{kh_K}{p_K} \leq C_1 \quad \forall K \in \mathcal{T}, \quad (3.27)$$

for some  $C_0, C_1 > 0$ , together with appropriate geometrical mesh refinement in neighbourhoods of the polygon vertices are sufficient to bound the adjoint approximation property  $\sigma_k^*(S_{\mathcal{T}}^p)$  (see [47, Thm. 2.4.2] and [36, 37]). The constant  $M_{\frac{kh}{p}}$  is then controlled by  $C_1$ . The above conditions are easily satisfied and imply that only  $O(1)$  degrees of freedom per wave length and per coordinate direction are necessary to obtain a  $k$ -independent reliability estimate.

- c. Note that in the reliability estimate the factor  $\sigma_k^*(S_{\mathcal{T}}^p) M_{\frac{kh}{p}}^{3/2}$  appears and in the efficiency estimate the factor  $M_{\frac{kh}{p}}$  appears. This indicates that for large  $M_{\frac{kh}{p}}$  the estimator might overestimate or underestimate the error, whereas a large value of  $\sigma_k^*(S_{\mathcal{T}}^p)$  suggests that the error might be underestimated (cf. [3, 29, 42] and also Fig. 2).

**Remark 3.9.** The proof of (3.25) implies that the jump term  $\eta_J$  in the error estimator can be omitted under two mild restrictions: a) The constant  $\mathfrak{a}$  in (2.8b) must satisfy  $\mathfrak{a} \geq C_{\mathfrak{a}} > 0$  for a sufficiently large constant  $C_{\mathfrak{a}}$  which only depends on the shape regularity via  $\rho_{\mathcal{T}}$ . However, explicit estimates for  $C_{\mathfrak{a}}$  are not available yet. b) The edge terms in the right-hand side of (3.25) are by a factor  $\sqrt{p_e}$  larger compared to edge residuals  $\eta_{E_K}$  and this leads to a reliability error estimate for the error estimator without jump term  $\eta_J$  which is polluted by a factor  $\sqrt{p_e}$ . However, the a priori analysis in [37] and [38] indicates that  $p \sim \log k$  is a typical choice so that this pollution is expected to be quite harmless.

## 4 Numerical Experiments

In this section we will report on numerical experiments to get insights in the following questions: a) How sharp does the error estimator reflect the behavior of the true error for uniform as well as for adaptive mesh refinement. b) How does the error estimator behave for scenarios which are not covered by our theory: for non-constant wavenumbers as well as for non-convex domains.

We have realized the dG-discretization with MATLAB and based the implementation on the finite element toolbox LehrFEM<sup>2</sup>.

The error in this section will be measured in the norm

$$\|u\|_{\mathcal{H};\mathcal{T}} := \|ku\| + \|\nabla_{\mathcal{T}} u\|.$$

---

<sup>2</sup><http://www.sam.math.ethz.ch/~hiptmair/tmp/LehrFEMManual.pdf>

## 4.1 Adaptive Algorithm

First, we will briefly describe our adaptive algorithm and refer for details, e.g., to [41]. It consists of the following flow of modules: **SOLVE**  $\longrightarrow$  **ESTIMATE**  $\longrightarrow$  **MARK**  $\longrightarrow$  **REFINE** and we will comment on their realization next.

### 4.1.1 Solve

The module **SOLVE** finds the solution  $u_{\mathcal{T}}$  of (2.8) for a given mesh  $\mathcal{T}$  with polynomial degree function  $\mathbf{p}$  and data  $f, g, k, \Omega$ . In our implementation all integrals involved in (2.8) are computed by quadrature on edges and elements.

### 4.1.2 Estimate

As explained in Remark 3.9 we have omitted the jump term  $\eta_J$  and realized the right-hand side in (3.25) as the error estimator. For simplicity we have also omitted the oscillation terms and worked with the functions  $f, g$  instead. Again, all integrals are computed via numerical quadrature. The resulting local and global error estimator are denoted by

$$\tilde{\eta}_K^2 := \eta_{R_K}^2 + \frac{p_K}{2} \left\| \sqrt{\frac{\mathbf{h}}{\mathbf{p}}} \llbracket \nabla_{\mathcal{T}} u_{\mathcal{T}} \rrbracket_N \right\|_{\partial^I K}^2 + \left\| \sqrt{\mathbf{h}} (g - (\partial_{\mathbf{n}} + \mathbf{i} k) u_{\mathcal{T}}) \right\|_{\partial^B K}^2$$

and

$$\tilde{\eta} := \sum_{K \in \mathcal{T}} \tilde{\eta}_K^2,$$

where the notation “ $\doteq$ ” indicates that the left-hand side equals the right-hand side up to numerical quadrature.

### 4.1.3 Mark

After having computed the local estimators  $\tilde{\eta}_K$  a refinement strategy has to be applied and we employ Dörfler’s marking strategy: Fix the triangulation  $\mathcal{T}$  and let  $u_{\mathcal{T}} \in S_{\mathcal{T}}^{\mathbf{p}}$  be the dG-solution. Denote by  $\mathcal{S}$  some subset of  $\mathcal{T}$ . We write

$$\tilde{\eta}(u_{\mathcal{T}}, \mathcal{S}) := \sum_{K \in \mathcal{S}} \tilde{\eta}_K^2(u_{\mathcal{T}}).$$

For fixed threshold  $\theta \in ]0, 1]$ , the set of marked elements  $\mathcal{M} \subseteq \mathcal{T}$  is defined by

$$\mathcal{M} := \operatorname{argmin} \{ \operatorname{card}(\mathcal{S}) \mid \mathcal{S} \subseteq \mathcal{T} \wedge \tilde{\eta}(u_{\mathcal{T}}, \mathcal{S}) \geq \theta \tilde{\eta}(u_{\mathcal{T}}, \mathcal{T}) \}.$$

### 4.1.4 Refine

In this step, all elements  $K \in \mathcal{M}$  are refined. Some additional elements are refined to eliminate hanging nodes and we have realized the largest edge bisection for this purpose. We emphasize that our implementation is currently restricted to  $h$  refinement while an extension to adaptive  $hp$ -refinement will be the topic of future research.



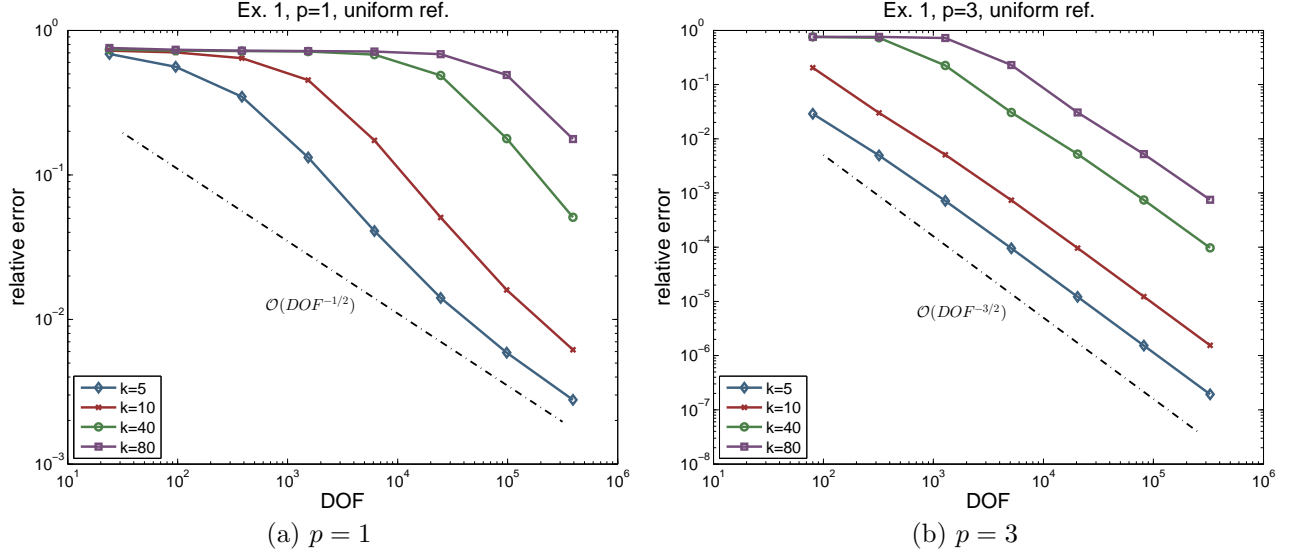


Figure 1: Comparison of the relative error in the norm  $\| \cdot \|_{\mathcal{H};\mathcal{T}}$ , for the polynomial degrees  $p = 1$  and  $p = 3$  for different values of  $k$  in Example 1.

## 4.2 Plane Wave Solutions

The parameters  $\mathbf{a} = 30$ ,  $\mathbf{b} = 1$ , and  $\mathbf{d} = 1/4$  in (2.8) are fixed for all experiments in this section. The adaptive refinement process is always started on a coarse mesh where the number of mesh cells is  $O(1)$  independent of  $k$  and  $p$ .

### 4.2.1 Example 1

Let  $\Omega = (0, 1)^2$  and the data  $f, g$  be given such that  $u(x, y) := \exp(i k (x + y))$  is the exact solution. As  $u$  is an entire function it is reasonable to refine the mesh uniformly. In Fig. 1, we compare the relative error in the  $\| \cdot \|_{\mathcal{H}}$  norm for different wavenumbers. As expected a) the pollution effect is visible, i.e., the convergence starts later for higher wavenumbers and b) the pollution becomes smaller for higher polynomial degree.

Next we test the sharpness of the reliability estimate for the error estimator. In Fig. 2 the ratio  $\|u - u_{\mathcal{T}}\|_{\mathcal{H};\mathcal{T}} / \check{\eta}(u_{\mathcal{T}})$  for different polynomial degrees and wavenumbers are depicted. Since we start with a very coarse initial mesh the constant  $M_{\frac{k\mathbf{h}}{p}}$  increases with increasing  $k$  in the pre-asymptotic regime and, due to Remark 3.8.c, an underestimating can be expected (as compared to when the asymptotic regime is reached). This effect can be seen in Fig. 2 while the asymptotic regime is reached faster for higher order polynomial degree.

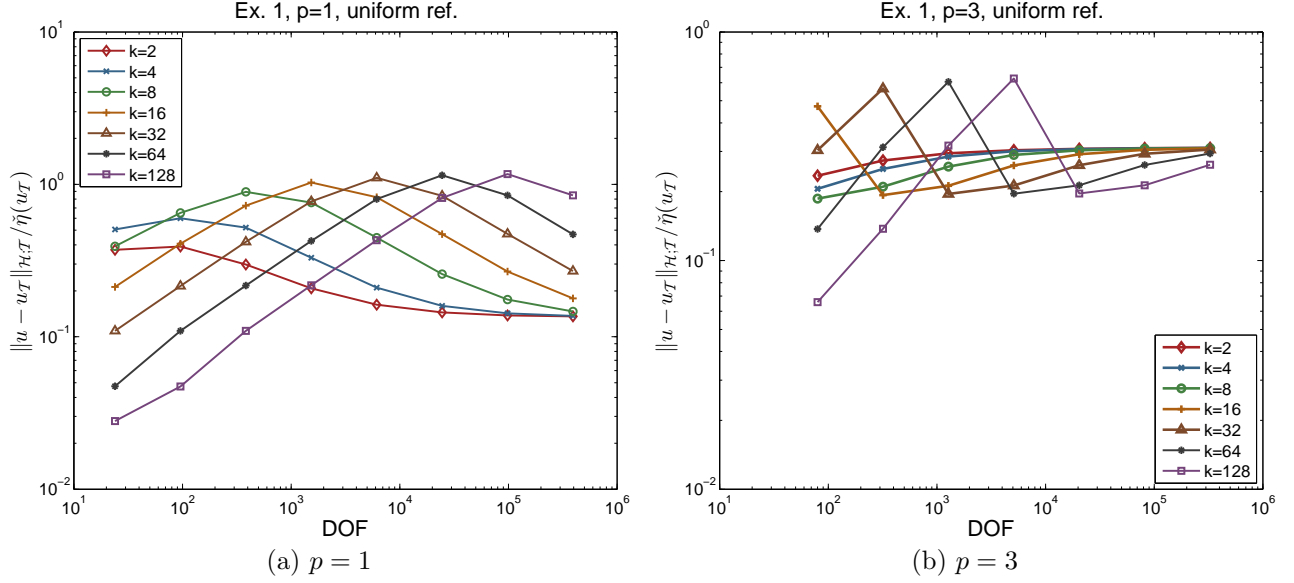


Figure 2: Ratio of the exact error  $\|u - u_{\mathcal{T}}\|_{\mathcal{H};\mathcal{T}}$  and the estimated error  $\tilde{\eta}(u_{\mathcal{T}})$  for different values of  $k$  in Example 1.

#### 4.2.2 Example 2

We consider the Helmholtz problem on  $\Omega = (0, 2\pi)^2$  with the exact solution  $u(x, y) = \exp(ikx)$ . The corresponding functions  $f$  and  $g$  are chosen accordingly:

$$f := 0 \quad \text{and} \quad g(x, y) := \begin{cases} 0 & \text{if } x = 0, \\ 2ik & \text{if } x = 2\pi, \\ ik e^{ikx} & \text{otherwise,} \end{cases} \quad \forall (x, y) \in \partial\Omega. \quad (4.1)$$

The dG-solution for very coarse meshes is strongly polluted and does not reflect the uniformly oscillating behavior, e.g., in the imaginary part  $\text{Im } u = \sin kx$  of the solution. One possible interpretation is that  $f = 0$  in  $\Omega$  and  $g = 0$  at the left boundary have the effect that  $u_{\mathcal{T}}$  is small close to the left boundary while at the right boundary the oscillations got resolved earlier. This is “seen” also by the error estimator and stronger refinement takes place in the early stage of adaptivity close to the right boundary. Only after some refinement steps the strong mesh refinement penetrates from right to left into the whole domain (see. Fig. 3). In Fig. 4(a), we see that the mesh starts to become uniform as soon as the resolution condition (3.27) is fulfilled and the error starts to decrease.

Furthermore we emphasize the following two points.

- a. As is well-known reliability is not a local property and we have here an example where the local error indicator  $\tilde{\eta}_K$  differs significantly from the local error in the left part of the domain in the pre-asymptotic regime. In addition,  $M_{\frac{kh}{p}}$  is large and due to Remark 3.8.c the underestimation of the error in this early stage of refinement can be explained. This behavior is illustrated in Fig. 4(b).

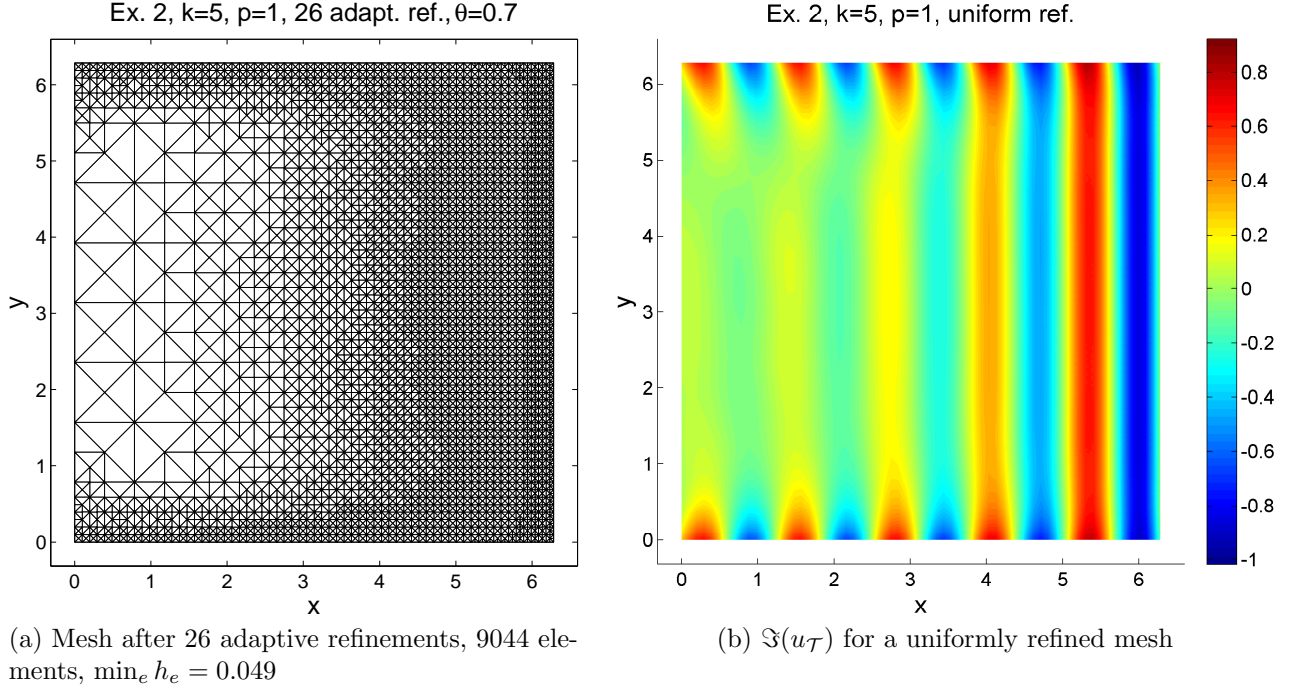


Figure 3: Adaptive mesh and imaginary part of the DGFEM solution for a uniform mesh with large mesh width,  $k = 5$ , and  $p = 1$  in Example 2. The exact solution is  $u(x, y) = \exp(ikx)$ , and therefore  $\Im(u(x, y)) = \sin(kx)$ .

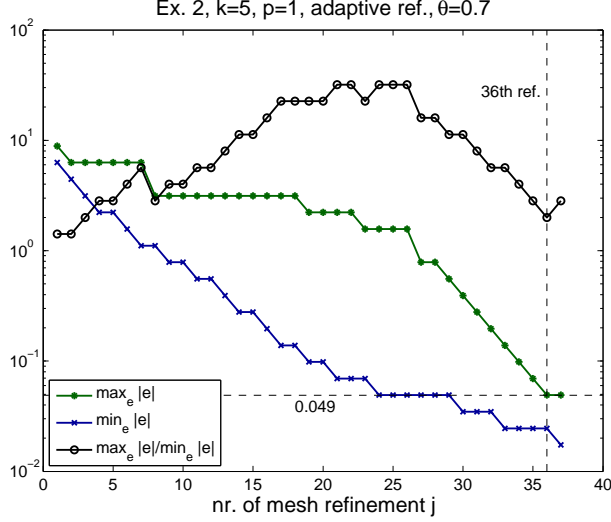
- b. It is also worth mentioning that we start the adaptive discretization with a very coarse initial mesh where the resolution condition (3.27) is not fulfilled for a moderate constant  $C_0$ . The numerical experiments indicate that the adaptive process behaves robustly for the dG-formulation already in the pre-asymptotic regime.

### 4.3 Example 3: L-shaped Domain

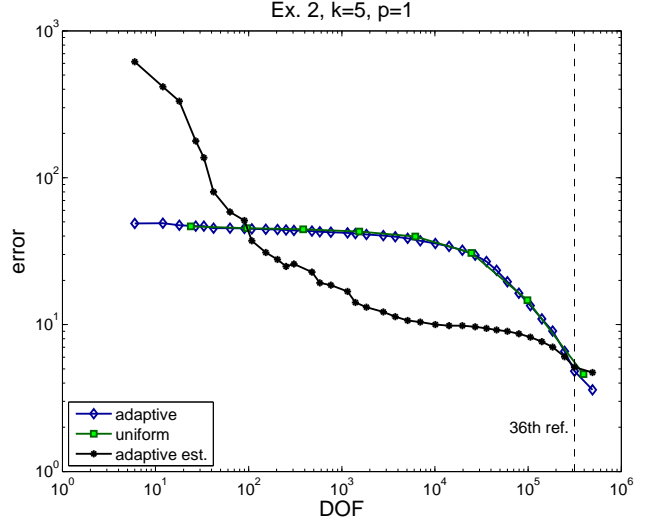
In this example we consider the L-shaped domain  $\Omega := (-1, 1)^2 \setminus ([0, 1] \times [-1, 0])$  with right-hand sides  $f$  and  $g$  chosen such that the first kind Bessel function  $u(x, y) := J_{1/2}(kr)$  with  $r := \sqrt{x^2 + y^2}$  is the exact solution (see also [26]). The Bessel function and solution  $u$  are plotted in Fig. 5. The problem is chosen such that the solution has a singularity at the reentrant corner located at  $\mathbf{0}$ .

In Fig. 6, two meshes generated by the adaptive procedure are depicted for uniform polynomial degree  $p = 1$  and wavenumber  $k = 10$ . The oscillating nature of the solution as well as the singular behavior is nicely reflected by the distribution of the mesh cells.

In Fig. 7, we compare uniform with adaptive mesh refinement for different values of  $k$  and  $p$ . As expected the uniform mesh refinement results in suboptimal convergence rates while the optimal convergence rates are preserved by adaptive refinement for the considered polynomial degrees  $p = 2, 4$ . In both cases some initial refinement steps are required before the error starts to decrease due to the pollution effect. Again the pollution is significantly reduced for



(a) In this plot  $|e|$  denotes the length of the edge  $e$ . The plot shows the maximum length of an edge, the minimum length of an edge, and the ratio for the  $j$ -th adaptively refined mesh.



(b) Error  $\|u - u_{\mathcal{T}}\|_{\mathcal{H};\mathcal{T}}$  for uniform and adaptive refinement (with  $\theta = 0.7$ ) and the estimated error  $\tilde{\eta}(u_{\mathcal{T}})$  for adaptive refinement

Figure 4: In Fig. (a) it can be seen that the adaptive algorithm, applied to Example 2 with  $k = 5$  and  $p = 1$ , at first generates a mesh with very diverse element sizes, which then turns into an almost uniform mesh at about the 36th refinement. This refinement corresponds to a maximum edge length of 0.049. In both plots, the dashed line marks this mesh width, respectively the point at which this adaptive refinement takes place. We observe that convergence for uniform refinement starts shortly before this mesh size is reached. Moreover, at this refinement, the error estimator surpasses the actual error in this example, and the error is underestimated in the preasymptotic range.

higher polynomial degree.

#### 4.4 Example 4: Non-constant Wavenumber

In this section, we consider the case of non-constant wavenumber  $k$  which has important practical applications. Although we have formulated the dG-method for non-constant wavenumber our theory only covers the constant case. Nonetheless the numerical experiments indicate that the a posteriori error estimation leads to an efficient adaptive solution method.

Consider the domain  $\Omega = (0, 2\pi)^2$ . We partition  $\Omega$  into the disc  $\Omega_1$  about  $(\pi, \pi)^\top$  with radius  $3/2$  and its complement  $\Omega_2 := \Omega \setminus \Omega_1$ . Let  $k_1, k_2 > 0$ . The function  $k$  is defined piecewise by  $k|_{\Omega_i} := k_i$ ,  $i = 1, 2$ . We have chosen  $f = 0$  and

$$g_1(x, y) := \begin{cases} -1 & x = 0, \\ i & x = 2\pi, \\ 0 & \text{otherwise,} \end{cases} \quad \forall (x, y) \in \partial\Omega. \quad (4.2)$$

Alternatively we will consider boundary data as defined in (4.1) with  $k := k_2$  and denote them

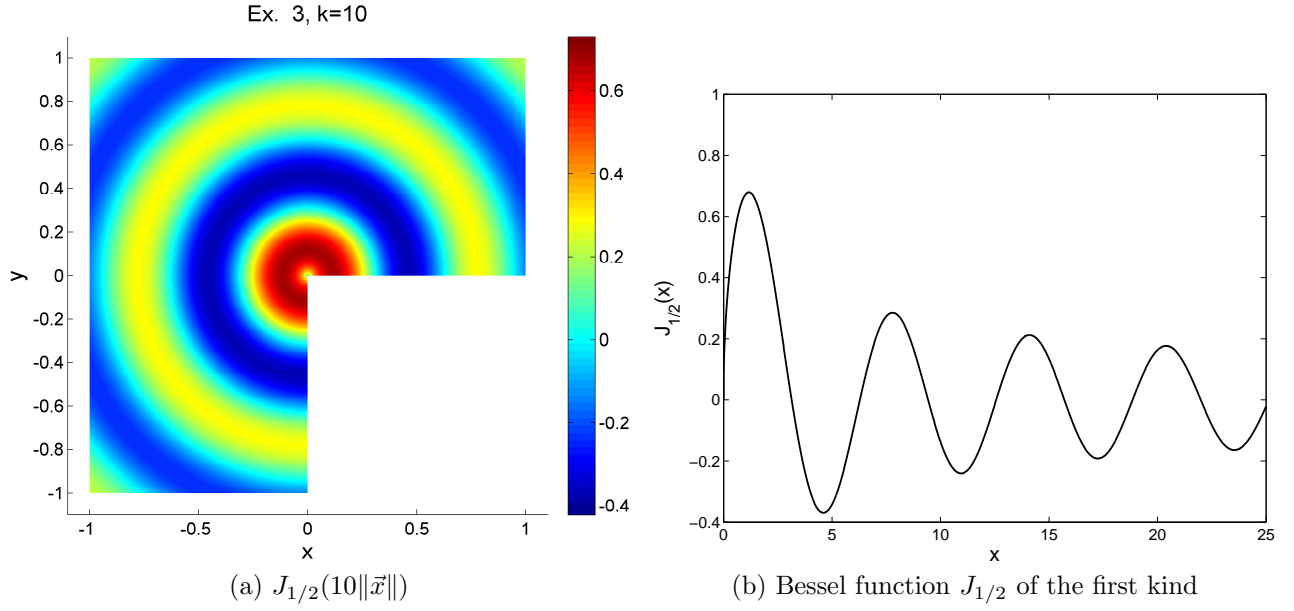


Figure 5: The solution  $u = J_{1/2}(kr)$  in Example 3 for  $k = 10$ , and the Bessel function  $J_{1/2}(x)$ , whose derivative goes to infinity for  $x \rightarrow 0$ .

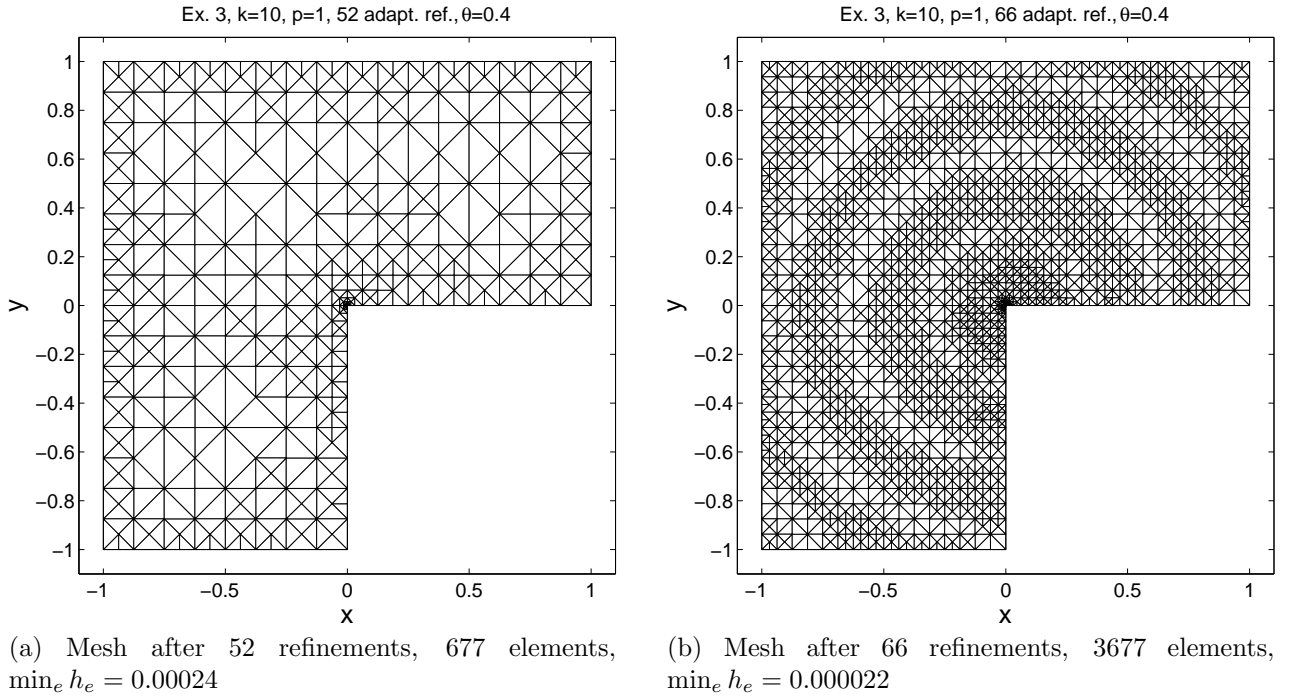


Figure 6: Meshes obtained by the adaptive algorithm for Example 3.

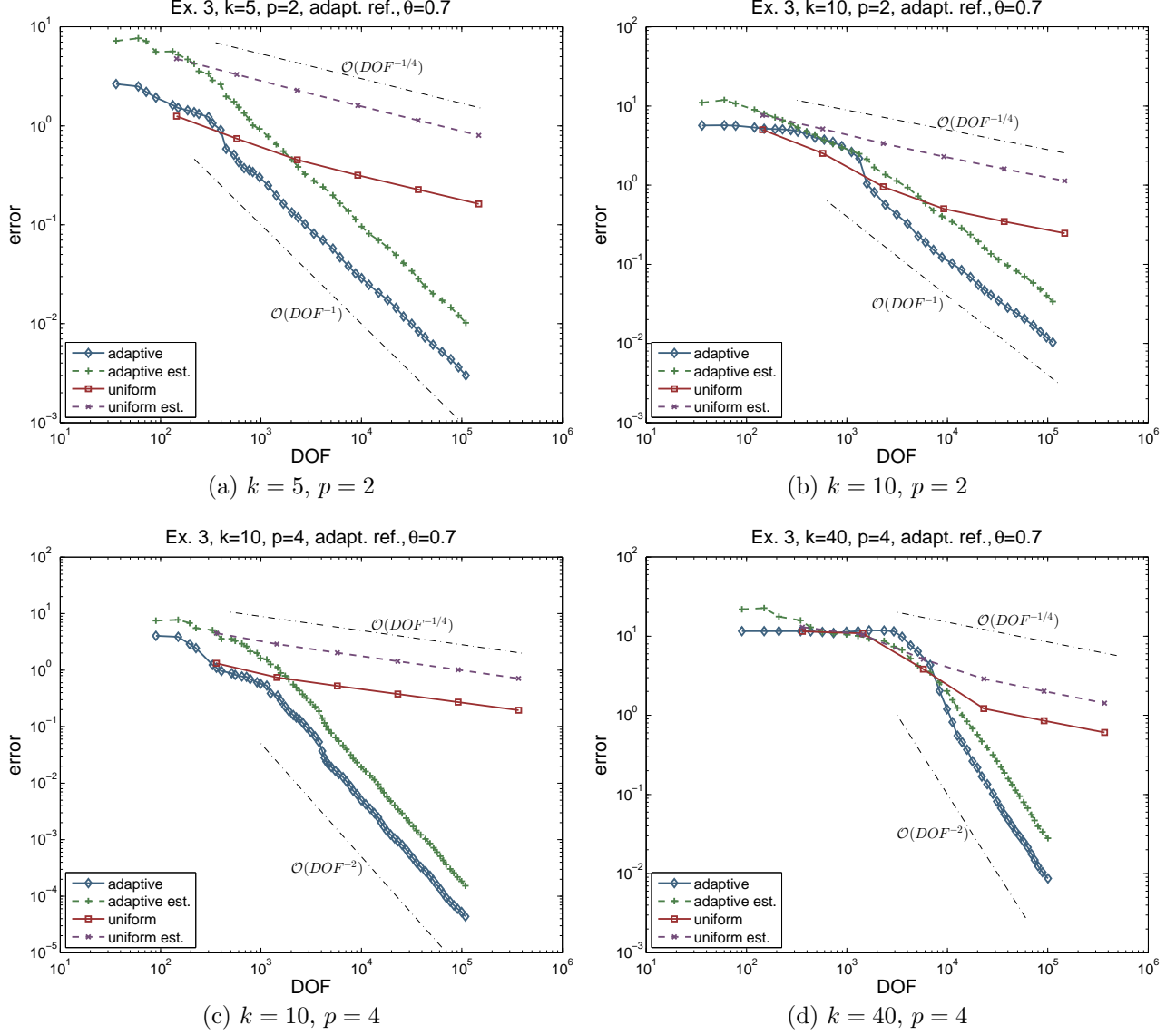


Figure 7: Comparison of the actual error  $\|u - u_{\mathcal{T}}\|_{\mathcal{H};\mathcal{T}}$  and the estimated error  $\check{\eta}(u_{\mathcal{T}})$ , using uniform and adaptive refinement with  $\theta = 0.7$  in Example 3 for different values of  $k$  and  $p$ .

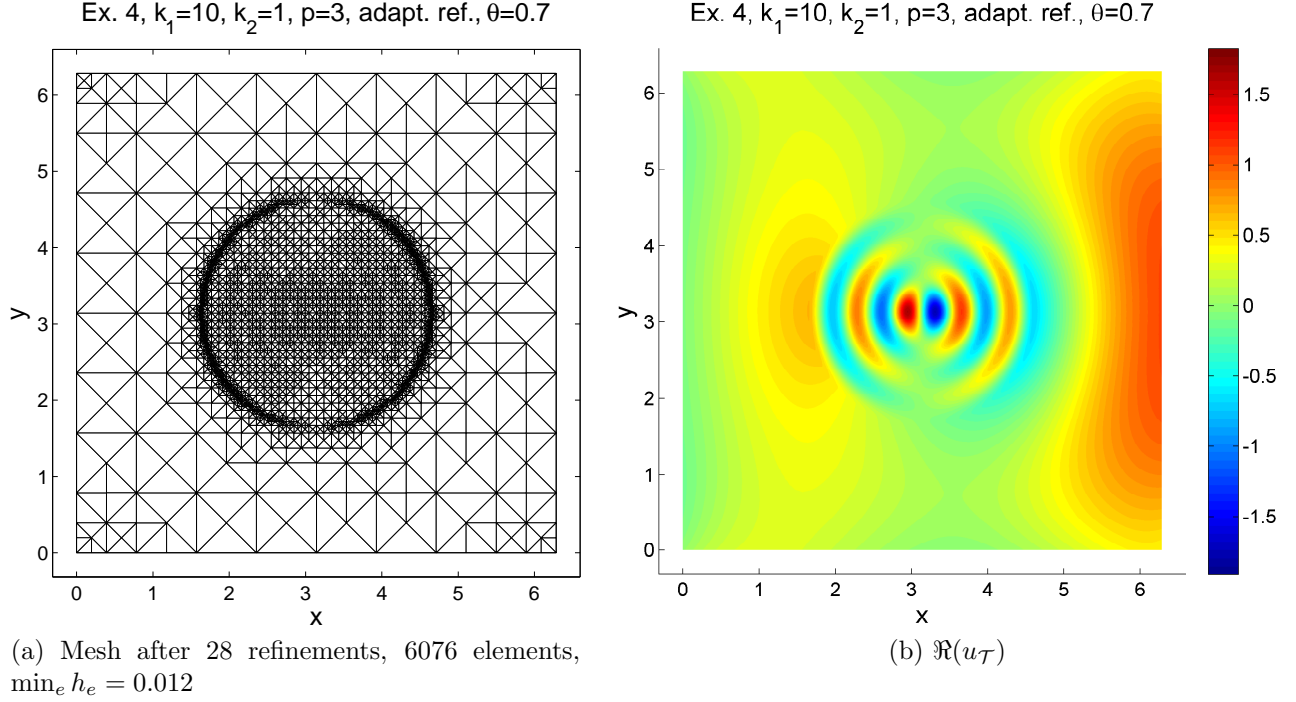


Figure 8: Adaptively refined mesh with  $\theta = 0.7$  and real part of the DGFEM solution on this mesh for Example 4 with  $k_1 = 10$ ,  $k_2 = 1$ , and the boundary data  $g_1$ .

here by  $g_2$ .

In Fig. 8, the adaptively refined mesh and the real part of the dG-solution are plotted for  $k_1 = 10$ ,  $k_2 = 1$ , and the boundary data  $g_1$ . Strong refinement takes place in the vicinity of the circular interface between  $\Omega_1$  and  $\Omega_2$ . Moreover, the mesh width is much smaller inside the circle, where the wavenumber is high in accordance with the smoothness properties of the solution.

In the next example we have considered the reversed situation:  $k_1 = 1$ ,  $k_2 = 10$ , and boundary data  $g_2$ . Fig. 9 implies, that strong refinement close to the jump of the wavenumber is not always necessary. In this case, the solution appears to be smooth, respectively almost zero near the left part of the inner circle where  $k = k_2$  holds and this is taken into account by the adaptive algorithm. Fig. 10 reflects the convergence of the estimated error for Example 4.

These examples indicate that the adaptive algorithm, applied with the error estimator  $\tilde{\eta}_K$ , properly accomplishes the task of refining the mesh according to the properties of the solution: Singularities and wave characteristics are recognized by the estimator, and we observed optimal convergence rates.



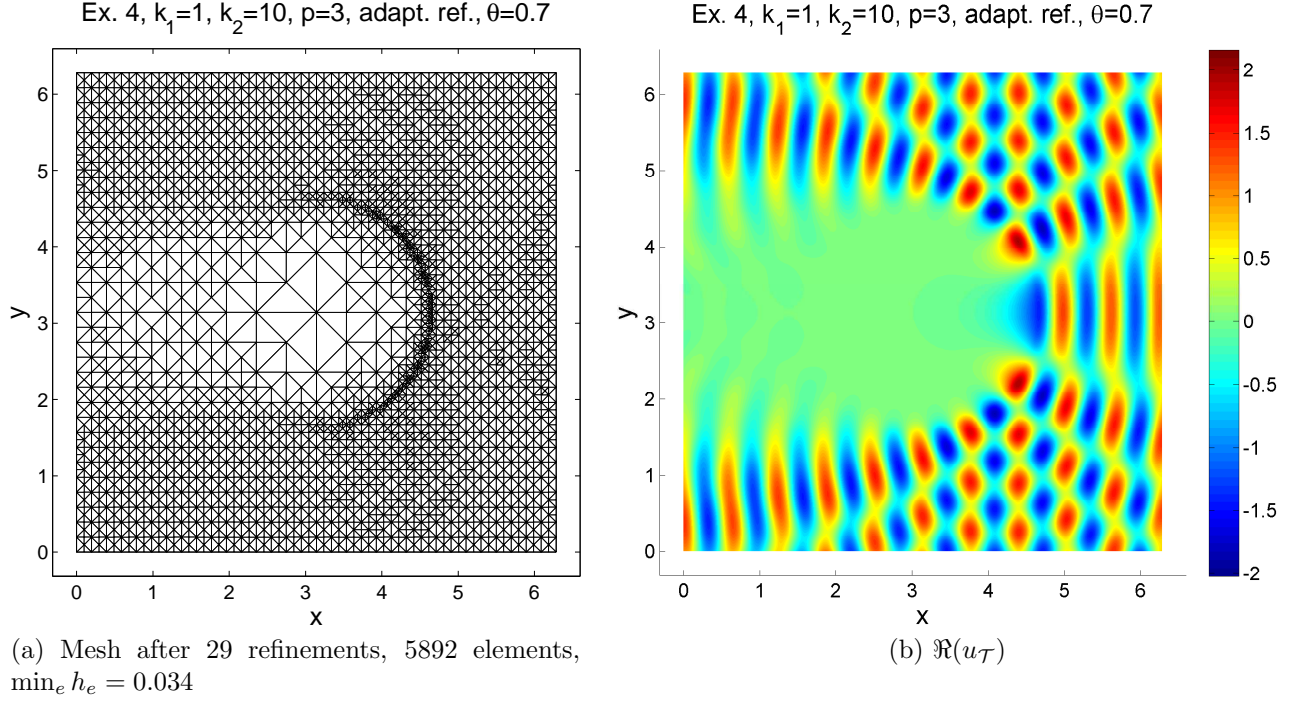


Figure 9: Adaptively refined mesh with  $\theta = 0.7$  and real part of the DGFEM solution on this mesh for Example 4 with  $k_1 = 1$ ,  $k_2 = 10$ , and the boundary data  $g_2$ .

## 5 Conclusion and Outlook

In this paper we have derived an a posteriori error estimator for an  $hp$ -dG method for highly indefinite Helmholtz problems. In contrast to the discretization of the standard variational formulation of the Helmholtz problem, the chosen  $hp$ -dG discretization always has a unique solution (cf. Remark 2.4). We have proved reliability and efficiency estimates which are explicit in the discretization parameters  $h$ ,  $p$ , and the wave number  $k$ . Note that the adjoint approximation property  $\sigma_k^*(S_{\mathcal{T}}^p)$  enters the reliability estimate. In [47, Thm. 2.4.2] and [36, 37] it has been proved that for convex polygonal domains the conditions

$$p \geq C_0 \log k \quad \text{and} \quad \forall K \in \mathcal{T} : \quad \frac{kh_K}{p_K} \leq C_1 \quad (5.1)$$

imply  $\sigma_k^*(S_{\mathcal{T}}^p) \leq C_2$ . We expect that general polygonal domains can be handled by a) generalizing the “decomposition lemma” [38, Theorem 4.10] to a weighted  $H^2(\Omega)$ -regularity estimate for the non-analytic part of the adjoint solution and b) performing an appropriate mesh grading towards reentrant corners as is well known for elliptic boundary value problem. Then, the error estimate for the non-analytic part can be derived in a similar fashion as the estimate of  $\eta_{\mathcal{A}}$  in the proof of [38, Proposition 5.6]. Again, we expect that the resolution condition (5.1) remains unchanged while the constant  $C_1$  then, possibly, depends on the angles at the reentrant corners of the polygon. Whereas the rigorous derivation of such estimates is a topic of future research, we point out that the use of adaptive methods is already justified in our



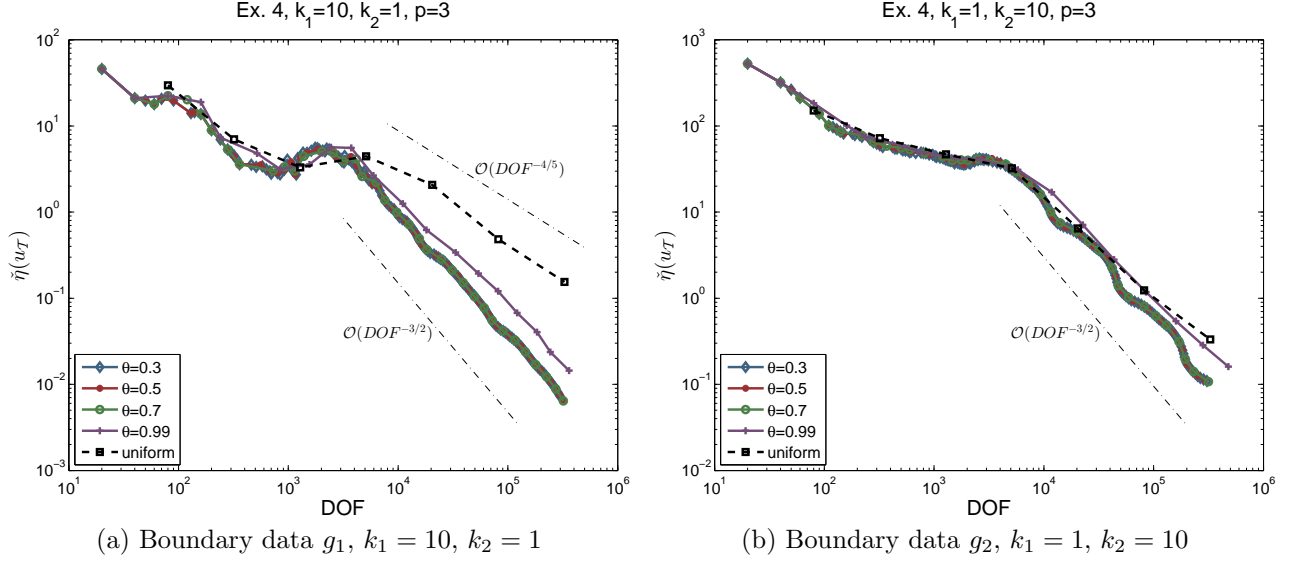


Figure 10: The convergence of the error estimator  $\tilde{\eta}(u_T)$  in Example 4 for two non-constant functions  $k(x, y)$  and the boundary data  $g_1, g_2$ , respectively.

model setting of convex polygons, since higher polynomial degrees require graded meshes also at convex corners in order to preserve optimal convergence rates.

Our analysis is not sharp enough to give precise bounds for the constants  $C_0, C_1, C_2$ . The numerical experiments show that these estimates are *qualitatively* sharp, i.e., if the polynomial degree stays fixed independent of  $k$ , the error estimator significantly overestimates the error while a mild, logarithmic increase depending on  $k$  cures this problem. It would be also interesting to estimate the size of  $\sigma_k^*(S_{\mathcal{T}}^p)$  by numerical experiments. However, this task is far from being trivial because the adjoint approximation property is defined as an infinite-dimensional sup-inf problem, and the dependence on the regularity of the domain, step size  $h$ , polynomial degree  $p$ , wave number  $k$  requires extensive numerical tests which would increase the length of the paper substantially. We are planning to investigate this question as a topic of further research. Our numerical examples indicate that, as soon as the resolution condition is satisfied with constants  $C_0 \sim C_1 = \mathcal{O}(1)$ , the a posteriori error estimator becomes quite sharp.

Another interesting question is related to the mesh grading towards the corners of the polygonal domain. The results in [38] imply that if the initial, coarsest mesh and polynomial degrees are chosen according to (5.1) and [38, Assumption 5.4] then,  $\sigma_k^*(S_{\mathcal{T}}^p)$  stays bounded by a constant during the whole adaptive process and the geometric grading may not to be incorporated in the adaptive refinement procedure. Our numerical experiments show that after some refinements (as soon as the resolution condition is satisfied) the convergence rate of the adaptive solution becomes optimal and, in addition, the error estimator nicely reflects the size and decay of the error. This behaviour of the estimator, which is supported by our analysis only in case  $\sigma_k^*(S_{\mathcal{T}}^p)$  that is moderate, suggests that the adaptive algorithm achieves an appropriate mesh grading on its own.

# A Approximation Properties

## A.1 $C^1$ -hp Interpolant

For residual-type a posteriori error estimation, typically, the subtle choice of an interpolation operator for the approximation of the error along  $hp$ -explicit error estimates plays an essential role. For our *non-conforming* dG-formulation it turns out that a  $C^1$ -interpolation operator has favorable properties, namely, the internal jumps vanish while the approximation estimates are preserved. In [39] a  $C^0$ -hp-Clément-type interpolation operator is constructed and  $hp$ -explicit error estimates are derived for  $W^{1,q}(\Omega)$  functions. In contrast, our estimate for the  $C^1$ -hp Clément interpolation operator allows for higher-order convergence estimates for smoother functions as well as for estimates in norms which are stronger than the  $H^1$ -norm. The proof follows the ideas in [39, Thm. 2.1] and employs a  $C^1$ -partition of unity by the quintic Argyris finite element.

The construction is in two steps. First local (discontinuous) approximations are constructed on local triangle patches. By multiplying with a  $C^1$ -partition of unity the resulting approximation is in  $C^1(\Omega)$ , while the approximation properties are preserved.

The first step is described by the following theorem. Its proof can be found in [35, Thm. 5.1] which is a generalization of the one-dimensional construction (see, e.g., [14, Chap. 7, eq. (2.8)]).

**Theorem A.1.** *Let  $d \in \mathbb{N}$  and  $I := \times_{j=1}^d I_j$  with  $I_i$  being a bounded interval for every  $i \in \{1, \dots, d\}$ . Let  $n \in \mathbb{N}$ . Then, for any  $p \in \mathbb{N}$  with  $p \geq n - 1$ , there exists a bounded linear operator  $J_{n,p} : L^1(I) \rightarrow \bigotimes_{j=1}^d \mathbb{P}_p(I_j)$  with the following properties: For each  $q \in [1, \infty]$ , there exists a constant  $C > 0$  depending only on  $n$ ,  $q$ , and  $I$  such that for all  $0 \leq n \leq N$*

$$J_{n,p}u = u \quad \forall u \in \bigotimes_{j=1}^d \mathbb{P}_{n-1}(I_j)$$

$$\|u - J_{n,p}\|_{W^{\ell,q}(I)} \leq C (N+1)^{-(r-\ell)} |u|_{W^{r,q}(I)}, \quad 0 \leq \ell \leq r \leq n.$$

The proof of the following theorem is a generalization of the proof of [39, Thm. 2.1] and is carried out in detail in [47, Thm. 3.1.10]. Here we skip it for brevity.

**Theorem A.2** (Clément type quasi-interpolation). *Let  $\mathcal{T}$  be a  $\rho_{\mathcal{T}}$ -shape regular, conforming simplicial finite element mesh for the polygonal Lipschitz domain  $\Omega \subseteq \mathbb{R}^2$ . Let  $\mathbf{p}$  be a polynomial degree function for  $\mathcal{T}$  satisfying (2.4). Assume that  $q \in [1, \infty]$  and let  $n \in \mathbb{N}$ .*

- a. *Assume that  $\lfloor (p_{\mathcal{T}} - 5)/2 \rfloor \geq n - 1$ . Then, there exists a bounded linear operator  $I_n^{\text{hp}} : W^{n,q}(\Omega) \rightarrow S_{\mathcal{T}}^{\mathbf{p}} \cap C^1(\Omega)$  such that for every  $K \in \mathcal{T}$*

$$|u - I_n^{\text{hp}}u|_{W^{m,q}(K)} \leq C \left( \frac{h_K}{p_K} \right)^{n-m} |u|_{W^{n,q}(\omega_K)} \quad \forall m \in \{0, \dots, n\}, \quad (\text{A.1a})$$

and for every  $e \in \mathcal{E}(K)$  and multiindex  $\vartheta \in \mathbb{N}_0^2$  with  $\vartheta_1 + \vartheta_2 = m \leq n - 1$

$$\left\| \frac{\partial^m}{\partial x^{\vartheta_1} \partial y^{\vartheta_2}} ((u - I_n^{\text{hp}}u)|_K) \right\|_{L^q(e)} \leq C \left( \frac{h_e}{p_e} \right)^{n-m-1/q} |u|_{W^{n,q}(\omega_e)}, \quad (\text{A.1b})$$

where  $C > 0$  only depends on  $n$ ,  $q$ ,  $\rho_{\mathcal{T}}$ , and  $\Omega$ .

- b. Assume that  $\lfloor (p_{\mathcal{T}} - 1) / 2 \rfloor \geq n - 1$ . Then, there exists a bounded linear operator  $I_n^{\text{hp},0} : W^{n,q}(\Omega) \rightarrow S_{\mathcal{T}}^{\mathbf{p}} \cap C^0(\Omega)$  such that (A.1) holds with  $I_n^{\text{hp}}u$  replaced by  $I_n^{\text{hp},0}u$  for a constant  $C > 0$  solely depending on  $n$ ,  $q$ ,  $\rho_{\mathcal{T}}$ , and  $\Omega$ .

## A.2 Conforming Approximation

The a posteriori error analysis for our non-conforming dG-formulation requires the construction of conforming approximants of non-conforming  $hp$ -finite element functions and this will be provided next.

**Theorem A.3** (Conforming approximant). *Let  $\mathcal{T}$  be a  $\rho_{\mathcal{T}}$ -shape regular, conforming simplicial finite element mesh of the polygonal domain  $\Omega \subseteq \mathbb{R}^2$ . Let  $v \in S_{\mathcal{T}}^{\mathbf{p}}$ , and let  $\mathbf{p}$  be a polynomial degree function satisfying (2.4) and  $p_{\mathcal{T}} \geq 1$ . Then, there exists a constant  $C > 0$  which only depends on the shape regularity and a function  $v^* \in S_{\mathcal{T}}^{\mathbf{p}} \cap C^0(\Omega)$  such that*

$$\|v - v^*\|_{\partial\Omega} \leq C \|\llbracket v \rrbracket\|_{\mathfrak{S}^I}, \quad (\text{A.2a})$$

$$\|v - v^*\| \leq C \|\mathfrak{h}^{1/2} \llbracket v \rrbracket\|_{\mathfrak{S}^I}, \quad (\text{A.2b})$$

$$\|\nabla(v - v^*)\| \leq C \left\| \frac{\mathbf{p}}{\mathfrak{h}^{1/2}} \llbracket v \rrbracket \right\|_{\mathfrak{S}^I}. \quad (\text{A.2c})$$

For the proof of this theorem we refer to [47, Thm. 3.2.7] (see also, e.g., [7, 27]).

**Corollary A.4** (Conforming error). *Let the assumptions of Theorem A.3 be satisfied. There exists a constant  $C > 0$  which only depends on the shape regularity constant  $\rho_{\mathcal{T}}$  such that, for every  $v \in S_{\mathcal{T}}^{\mathbf{p}}$ , there is a function  $v^* \in S_{\mathcal{T}}^{\mathbf{p}} \cap C^0(\Omega)$  with*

$$\begin{aligned} \|k(v - v^*)\|^2 + \|\nabla(v - v^*)\|^2 + \left\| \sqrt{k}(v - v^*) \right\|_{\partial\Omega}^2 \\ \leq \frac{C}{\mathbf{a}} \left( 1 + \frac{1}{p_{\mathcal{T}}} M_{\frac{\text{kh}}{\mathbf{p}}} + M_{\frac{\text{kh}}{\mathbf{p}}}^2 \right) \left\| \sqrt{\mathbf{a} \frac{\mathbf{p}^2}{\mathfrak{h}}} \llbracket v \rrbracket \right\|_{\mathfrak{S}^I}^2. \end{aligned}$$

*Proof.* The estimate follows by (A.2). □

## References

- [1] J. D. Achenbach. *Wave propagation in elastic solids*. North Holland, Amsterdam, 2005.
- [2] M. Ainsworth and J. T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*. Wiley, 2000.
- [3] I. Babuška, F. Ihlenburg, T. Strouboulis, and S. K. Gangaraj. A posteriori error estimation for finite element solutions of Helmholtz' equation I. The quality of local indicators and estimators. *Internat. J. Numer. Methods Engrg.*, 40(18):3443–3462, 1997.

- [4] I. Babuška, F. Ihlenburg, T. Strouboulis, and S. K. Gangaraj. A posteriori error estimation for finite element solutions of Helmholtz' equation II. Estimation of the pollution error. *Internat. J. Numer. Methods Engrg.*, 40(21):3883–3900, 1997.
- [5] I. Babuška and W. C. Rheinboldt. A-posteriori error estimates for the finite element method. *Internat. J. Numer. Meth. Engrg.*, 12:1597–1615, 1978.
- [6] I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.*, 15:736–754, 1978.
- [7] A. Bonito and R. H. Nochetto. Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method. *SIAM J. Numer. Anal.*, 48(2):734–771, 2010.
- [8] A. Buffa and P. Monk. Error estimates for the ultra weak variational formulation of the Helmholtz equation. *M2AN Math. Model. Numer. Anal.*, 42(6):925–940, 2008.
- [9] O. Cessenat and B. Després. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz equation. *SIAM J. Numer. Anal.*, 35:255–299, 1998.
- [10] O. Cessenat and B. Després. Using plane waves as base functions for solving time harmonic equations with the ultra weak variational formulation. *J. Computational Acoustics*, 11:227–238, 2003.
- [11] S. N. Chandler-Wilde, I. G. Graham, S. Langdon, and E. A. Spence. Numerical-asymptotic boundary integral methods in high-frequency acoustic scattering. *Acta Numer.*, 21:89–305, 2012.
- [12] P. Cummings and X. Feng. Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations. *Math. Models Methods Appl. Sci.*, 16(1):139–160, 2006.
- [13] B. Després. Sur une formulation variationnelle de type ultra-faible. *C. R. Acad. Sci. Paris Sér. I Math.*, 318(10):939–944, 1994.
- [14] R. A. DeVore and G. G. Lorentz. *Constructive Approximation*. Springer-Verlag, New York, 1993.
- [15] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*. Springer, Heidelberg, 2012.
- [16] W. Dörfler and S. Sauter. A posteriori error estimation for highly indefinite Helmholtz problems. *Comput. Methods Appl. Math.*, 13(3):333–347, 2013.
- [17] S. Esterhazy and J. M. Melenk. On stability of discretizations of the Helmholtz equation. In I. Graham, T. Hou, O. Lakkis, and R. Scheichl, editors, *Numerical Analysis of Multiscale Problems*, volume 83 of *Lect. Notes Comput. Sci. Eng.*, pages 285–324. Springer, Berlin, 2012.

- [18] X. Feng and H. Wu. Discontinuous Galerkin methods for the Helmholtz equation with large wave number. *SIAM J. Numer. Anal.*, 47(4):2872–2896, 2009.
- [19] X. Feng and H. Wu. *hp*-discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Math. Comp.*, 80(276):1997–2024, 2011.
- [20] X. Feng and Y. Xing. Absolutely stable local discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Math. Comp.*, 82(283):1269–1296, 2013.
- [21] C. Gittelsohn, R. Hiptmair, and I. Perugia. Plane wave discontinuous Galerkin methods: Analysis of the h-version. *Int. J. Numer. Meth. Engrg.*, 43(2):297–331, 2009.
- [22] M. Grigoroscuta-Strugaru, M. Amara, H. Calandra, and R. Djellouli. A modified discontinuous Galerkin method for solving efficiently Helmholtz problems. *Commun. Comput. Phys.*, 11(2):335–350, 2012.
- [23] I. Harari. A survey of finite element methods for time-harmonic acoustics. *Comput. Methods Appl. Mech. Engrg.*, 195(13-16):1594–1607, 2006.
- [24] I. Harari and T. J. R. Hughes. Galerkin/least-squares finite element methods for the reduced wave equation with nonreflecting boundary conditions in unbounded domains. *Comput. Methods Appl. Mech. Engrg.*, 98(3):411–454, 1992.
- [25] R. Hiptmair, A. Moiola, and I. Perugia. Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the *p*-version. *SIAM J. Numer. Anal.*, 49(1):264–284, 2011.
- [26] R. H. W. Hoppe and N. Sharma. Convergence analysis of an adaptive interior penalty discontinuous Galerkin method for the Helmholtz equation. *IMA J. Numer. Anal.*, 33(3):898–921, 2013.
- [27] P. Houston, D. Schötzau, and T. P. Wihler. Energy norm a posteriori error estimation of *hp*-adaptive discontinuous Galerkin methods for elliptic problems. *Math. Models Methods Appl. Sci.*, 17(1):33–62, 2007.
- [28] T. Huttunen and P. Monk. The use of plane waves to approximate wave propagation in anisotropic media. *J. Computational Mathematics*, 25:350–367, 2007.
- [29] F. Ihlenburg. *Finite Element Analysis of Acousting Scattering*. Springer, New York, 1998.
- [30] F. Ihlenburg and I. Babuška. Finite Element Solution to the Helmholtz Equation with High Wave Number. Part II: The h-p version of the FEM. *SIAM J. Num. Anal.*, 34(1):315–358, 1997.
- [31] S. Irinié and P. Bouillard. A residual a posteriori error estimator for the finite element solution of the Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 190(31):4027–4042, 2001.

- [32] J. D. Jackson. *Classical electrodynamics*. John Wiley & Sons, Inc., New York-London-Sydney, second edition, 1975.
- [33] R. Kellogg. Interpolation between subspaces of a Hilbert space. Technical Report BN-719, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland at College Park, College Park, MD, 20742-2431, USA, 1971.
- [34] J. M. Melenk. *On Generalized Finite Element Methods*. PhD thesis, University of Maryland at College Park, 1995.
- [35] J. M. Melenk.  $hp$ -interpolation of nonsmooth functions and an application to  $hp$ -a posteriori error estimation. *SIAM J. Numer. Anal.*, 43(1):127–155, 2005.
- [36] J. M. Melenk, A. Parsania, and S. Sauter. Generalized DG Methods for Highly Indefinite Helmholtz Problems Based on the Ultra-Weak Variational Formulation. *J Sci Comput*, 57:536–581, 2013.
- [37] J. M. Melenk and S. Sauter. Convergence Analysis for Finite Element Discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary condition. *Math. Comp*, 79:1871–1914, 2010.
- [38] J. M. Melenk and S. Sauter. Wave-Number Explicit Convergence Analysis for Galerkin Discretizations of the Helmholtz Equation. *SIAM J. Numer. Anal.*, 49(3):1210–1243, 2011.
- [39] J. M. Melenk and B. I. Wohlmuth. On residual-based a posteriori error estimation in  $hp$ -FEM. *Adv. Comput. Math.*, 15(1-4):311–331 (2002), 2001.
- [40] P. Monk and D.-Q. Wang. A least-squares method for the Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 175(1-2):121–136, 1999.
- [41] R. H. Nochetto, K. G. Siebert, and A. Veiser. Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation*, pages 409–542. Springer, Berlin, 2009.
- [42] J. T. Oden, S. Prudhomme, and L. Demkowicz. A posteriori error estimation for acoustic wave propagation problems. *Arch. Comput. Methods Engrg.*, 12(4):343–389, 2005.
- [43] A. H. Schatz. An observation concerning Ritz-Galerkin methods with indefinite bilinear forms. *Math. Comp.*, 28:959–962, 1974.
- [44] C. Schwab.  *$p$ - and  $hp$ -finite element methods*. The Clarendon Press Oxford University Press, New York, 1998.
- [45] R. Verfürth. *A posteriori error estimation techniques for finite element methods*. Oxford University Press, Oxford, 2013.
- [46] H. Wu. Pre-asymptotic error analysis of cip-fem and fem for the helmholtz equation with high wave number. part i: linear version. *IMA Journal of Numerical Analysis*, 2013.

- [47] J. Zech. A Posteriori Error Estimation of  $hp$ -DG Finite Element Methods for Highly Indefinite Helmholtz Problems. Master's thesis, Inst. f. Mathematik, Universität Zürich, 2014. <https://www.math.uzh.ch/compmath/index.php?id=zech>.
- [48] L. Zhu and H. Wu. Preasymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part II:  $hp$  version. *SIAM J. Numer. Anal.*, 51(3):1828–1852, 2013.